COGNITIVE STYLE AND ACHIEVEMENT: STUDENT PERFORMANCE ON THE THINKING STYLES INVENTORY, RAVEN'S PROGRESSIVE MATRICES, IOWA TEST OF BASIC SKILLS, AND THE FLORIDA WRITES! TEST

By

BRILEY E. PROCTOR

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1999

ACKNOWLEDGMENTS

As graduation day approaches, I reflect on the many, many people who have made this accomplishment possible. First and foremost, my committee chairperson, Dr. John Kranzler, deserves a special thanks for his mentorship over the past five years. Dr. Kranzler provided me with an abundance of guidance, research opportunities, teaching experiences, and professional advice that has made an indelible mark on my career. He epitomizes what all graduate students hope for in an academic mentor, and I feel extremely fortunate to have worked with and learned from such a fantastic professor. In addition, I am indebted to two other faculty members, Dr. Linda Crocker and Dr. James Algina, for their exceptional teaching in the area of educational measurement and evaluation. They set an example of high standards and thorough instruction that is simply unparalleled. I will always carry with me Dr. Crocker's resounding words of wisdom that she uttered quite frequently over these past few weeks - "Now Briley, don't panic!". I would also like to acknowledge Dr.s Mary K. Dykes, Edward W. Wolfe, and Jennifer Asmus for their continued support and guidance through both the qualifying exams and writing of the dissertation. All graduate students should have the type of committee experiences that I have had.

Outside of faculty members, there have been countless others who have provided me with infinite amounts of encouragement: to Bill and C.J., two of the best friends a girl

could ever ask for; Melanie, with whom a working relationship grew into a cherished friendship; my childhood pen-pal and long-time confidante, Anne, who kept me well-informed of "life on the outside"; Jim Jones, whose friendly bantering about the UF-FSU rivalry reminded me of what is really important in life. Perhaps most significantly, I would like to acknowledge my family: my parents, Proc and Becky, brothers, Edward and David, sister, Tracy, and each of their respective spouses. I was truly blessed to be given such a beautiful family, and I love each of them dearly.

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS ... ii

LIST OF TABLES ... vii

ABSTRACT ... xi

CHAPTERS

1 INTRODUCTION AND LITERATURE REVIEW ... 1

Performance Assessments ... 1
Definition and History ... 1
Performance versus Traditional Tests ... 3
Current State of Performance Assessments ... 7
Variables Affecting Student Performance on Tests ... 8
Gender and Race ... 8
Cognitive Ability, Personality, and Style ... 10
Previous Research on Styles ... 13
Cognition-Centered Styles ... 13
Personality-Centered Styles ... 20
Learning Styles ... 25
Sternberg's Cognitive Styles Theory ... 33
Theory of Mental Self-Government ... 34
Stability of Cognitive Styles ... 37
Cognitive Styles and School Success ... 37
Limitations of Studies on Cognitive Style and School Success ... 44
Statement of the Problem ... 45

page

2 METHODS ... 50

Participants ... 50
Instruments ... 50
Thinking Style Inventory (TSI) ... 50
Raven's Progressive Matrices (SPM) ... 55
Iowa Test of Basic Skills ... 57
FLORIDA WRITES! Assessment ... 59
Procedure ... 64
Analysis ... 66

3 RESULTS ... 67

Review of Analyses ... 67
Internal Consistency ... 71
Multiple Regression ... 72
Review of Procedure for Testing Main Hypotheses ... 72
Assumptions ... 73
Findings ... 74

4 DISCUSSION ... 92

Summary of Findings ... 92
Factor Analysis ... 93
Purpose ... 93
Results of Factor Analysis ... 95
Explanation of Subscales by Factor Models ... 101
Student Profiles Constructed from Factor Models ... 101
Summary of Factor Analysis Findings ... 104
Subscale Reliabilities ... 107
Implications for Research on the Construct of Style ... 108
Implications for Future Research on Performance Assessments ... 111
Limitations of the Current Study ... 112

APPENDICES

A CORRELATIONS OF ABILITIES, ACADEMIC PERFORMANCE, AND STYLES ... 117

B THINKING STYLE INVENTORY AND RESPONSE SHEET ... 118

page

C TESTING TIMES AND ITEM COUNTS FOR IOWA TEST OF BASIC SKILLS 125

D SCORING RUBRIC FOR FLORIDA WRITES! 126

E EIGHTH-GRADE SCORING RUBRIC FOR FLORIDA WRITES! 128

F SCORING DISTRIBUTIONS FOR FLORIDA WRITES! PROMPTS 131

G PARENT CONSENT FORM 133

H STUDENT ASSENT SCRIPT 135

I FULL CORRELATION MATRIX 137

REFERENCES 138

BIOGRAPHICAL SKETCH 146

LIST OF TABLES

Table	page
1. Correlation Coefficients of Thinking Style Inventory Subscale Scores with the Meyers-Briggs Type Indicator	52
2. Correlation Coefficients of the Thinking Style Inventory with the Gregorc Style Delineator	53
3. Interrater Reliability by Prompt for the Eighth-Grade Scores in 1995	64
4. Descriptive Statistics for Tests of Achievement and Cognitive Ability	67
5. Achievement and Ability Test Scores for Boys and Girls	69
6. Descriptive Statistics for Thinking Styles Inventory Subscales	70
7. Cronbach's Alpha (α) Estimates for the Thinking Styles Inventory Subscales (N=153)	71
8. Summary ANOVA Table for Reduced Model: Cognitive Ability (SPM) Regressed on Iowa Test of Basic Skills	75
9. Summary ANOVA Table for Full Model: Cognitive Ability (SPM) and Function Subscales (Executive, Legislative, Judicial) Regressed on Iowa Test of Basic Skills	76
10. Results: Cognitive Ability (SPM) and Function (Executive, Legislative, Judicial) Subscales Regressed on Iowa Test of Basic Skills	76
11. Summary ANOVA Table for Full Model: Cognitive Ability (SPM) and Level (Global, Local) Subscales Regressed on Iowa Test of Basic Skills	78
12. Results: Cognitive Ability (SPM) and Level (Global, Local) Subscales Regressed on Iowa Test of Basic Skills	79

Table page

13. Summary ANOVA Table for Full Model: Cognitive Ability (SPM) and Leaning (Liberal, Conservative) Subscales Regressed on Iowa Test of Basic Skills 79

14. Results: Cognitive Ability (SPM) and Leaning (Liberal, Conservative) Subscales Regressed on Iowa Test of Basic Skills 80

15. Summary ANOVA Table for Full Model: Cognitive Ability (SPM) and Form (Hierarchic, Monarchic, Oligarchic, Anarchic) Subscales Regressed on Iowa Test of Basic Skills 80

16. Results: Cognitive Ability (SPM) and Form (Hierarchic, Monarchic, Oligarchic, Anarchic) Subscales Regressed on Iowa Test of Basic Skills 81

17. Summary ANOVA Table for Full Model: Cognitive Ability (SPM) and Scope (Internal, External) Subscales Regressed on Iowa Test of Basic Skills 81

18. Results: Cognitive Ability (SPM) and Scope (Internal, External) Subscales Regressed on Iowa Test of Basic Skills 82

19. Summary ANOVA Table for Reduced Model: Cognitive Ability (SPM) and Iowa Test of Basic Skills Regressed on FLORIDA WRITES! 83

20. Results of Reduced Model: Cognitive Ability (SPM) and Iowa Test of Basic Skills Regressed on FLORIDA WRITES! 83

21. Summary ANOVA Table for Full Model: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Function (Executive, Legislative, Judicial) Regressed on FLORIDA WRITES! 84

22. Results: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Function (Executive, Legislative, Judicial) Subscales Regressed on FLORIDA WRITES! 85

23. Summary ANOVA Table for Full Model: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Level (Global, Local) Subscales Regressed on FLORIDA WRITES! 85

24. Results: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Level (Global, Local) Subscales Regressed on FLORIDA WRITES! 86

Table page

25. Summary ANOVA Table for Full Model: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Leaning (Liberal, Conservative) Subscales Regressed on FLORIDA WRITES! 86

26. Results: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Leaning (Liberal, Conservative) Subscales Regressed on FLORIDA WRITES! 87

27. Summary ANOVA Table for Full Model: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Form (Hierarchic, Monarchic, Oligarchic, Anarchic) Subscales Regressed on FLORIDA WRITES! 87

28. Results: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Form (Hierarchic, Monarchic, Oligarchic, Anarchic) Regressed on FLORIDA WRITES! 88

29. Summary ANOVA Table for Full Model: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Scope (Internal, External) Subscales Regressed on FLORIDA WRITES! 88

30. Results: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Scope (Internal, External) Regressed on FLORIDA WRITES! 89

31. Summary ANOVA Table for Reduced Model: Cognitive Ability (SPM) Regressed on FLORIDA WRITES! 90

32. Results of Reduced Model: Cognitive Ability (SPM) Regressed on FLORIDA WRITES! 90

33. Summary ANOVA Table for Full Model: Cognitive Ability (SPM) and Iowa Test of Basic Skills Regressed on FLORIDA WRITES! 91

34. Results of Full Model: Cognitive Ability (SPM) and Iowa Test of Basic Skills Regressed on FLORIDA WRITES! 91

35. Interscale Pearson Correlation Matrix for 13 Subscales of the Thinking Styles Inventory (N=153) 96

36. Factor 1 Loadings for the Unrotated Factor Solution 97

Table page

37. Varimax-Rotated Two-Factor Model for Thinking Styles Inventory 98

38. Varimax-Rotated Three-Factor Model for Thinking Styles Inventory 100

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

COGNITIVE STYLE AND ACHIEVEMENT: STUDENT PERFORMANCE ON THE THINKING STYLES INVENTORY, RAVEN'S PROGRESSIVE MATRICES, IOWA TEST OF BASIC SKILLS, AND THE FLORIDA WRITES! TEST

By

Briley E. Proctor

August 1999

Chairman: John H. Kranzler
Major Department: Foundations of Education

The purpose of this study was to investigate whether students' cognitive styles affect performance on two distinct types of tests: a standardized, multiple-choice test of language skills (Iowa Test of Basic Skills; ITBS), and a state-developed performance assessment of writing skill (FLORIDA WRITES!). The measure of cognitive style was the Thinking Style Inventory (TSI; Sternberg & Wagner, 1991), a self-report instrument based on Sternberg's (1997b) theory of thinking styles. Participants were 154 8th and 9th grade students. The main hypothesis was that cognitive style can explain variance in ITBS and FLORIDA WRITES! scores that is not accounted for by cognitive ability alone. Participants were administered the TSI and a measure of general cognitive ability, the Raven's Standard Progressive Matrices (SPM; Raven, Court, & Raven, 1983). ITBS and

FLORIDA WRITES! scores were obtained from the county data base. Multiple regression analyses were conducted, with thinking style and SPM as independent variables, and ITBS and FLORIDA WRITES! as dependent variables. Results of multiple regression analyses demonstrated that SPM was the only variable to correlate significantly with either ITBS or FLORIDA WRITES!. Cognitive style was not found to correlate with either dependent measure. An exploratory factor analysis of the TSI performed at the subscale level did not support Sternberg's five-factor structure; rather, a two-factor model was the best fit for the data. Some relationships between subscales were consistent with Sternberg's theory, whereas others were not supported. Limitations of the current study are discussed, in addition to implications for further research on the TSI. Finally, the relevance of cognitive style to students' academic success is considered.

CHAPTER 1
INTRODUCTION AND LITERATURE REVIEW

Performance Assessments

Definition and History

Although widely used within the educational milieu, the term "performance assessment" is so broad that it remains somewhat amorphous to many of its consumers. In the most simplistic terms, performance assessments are those that require examinees to demonstrate their mastery of skills and competencies by either performing or producing something (Finch, 1991). Examples of performance assessments in the field of education include giving music recitals, designing and conducting experiments, building models, giving speeches, and writing papers. Outside of education, the designation of performance assessment may apply to such diverse activities as contests measuring body building, floral arranging, and cake decorating, tests of marksmanship in the military, driver's license examinations, and Olympic figure-skating competitions.

Although performance assessment is currently receiving a great deal of attention in education and psychometrics, it is not a new concept. The earliest recorded measures of human achievement were, in fact, performance assessments: the civil service examinations of China (circa 2200 B.C.) required candidates to demonstrate competency in the areas of music, archery, horsemanship, writing, arithmetic, rituals, and ceremonies (DuBois, 1970;

Mueller, 1991). Modern forms of performance assessment may look different than those exemplified by the early Chinese civil service examinations, but they are predicated on similar theoretical foundations. Specifically, both the Chinese examinations and today's performance assessments are based on the premise that in order to determine one's mastery of a skill or competency, it is necessary to invoke and evaluate the specific skill or competency of interest. Although this assumption may appear unexceptionable, self-evident, and rather unimaginative, close examination of current testing practices reveals that the reality of today's educational assessments contrasts sharply with the aforementioned assumption (Resnick & Resnick, 1991).

The tests most frequently used in today's educational systems often do not directly assess the skill or competency of interest--rather, they assess indicators of the desired skill or competency (Resnick & Resnick, 1991). For example, to evaluate the ability to comprehend a written passage, some tests (e.g., Iowa Test of Basic Skills) require students to answer a series of multiple-choice items about some form of written prose. Although this type of test may correlate very highly with the construct of interest (the ability to comprehend written passages), it still remains an indicator of the ability, not a direct measure (Resnick & Resnick, 1991). In contrast, in a direct test, "the cognitive skill that is of interest is directly evaluated as it is expressed in the performance of some extended task" (Frederiksen & Collins, 1989, p. 28). Thus, a direct measure of reading comprehension would require observation of the examinee engaging in activities that demonstrate comprehension, such as describing the theme of the passage in her own words, elaborating on the presented points, and asking insightful questions. Another

example may be seen in the study of science. For instance, an essential component for the comprehension and interpretation of science is a solid understanding of the scientific method. To assess understanding of the scientific method, a teacher might administer a test that asks students to differentiate between examples and non-examples of the scientific method, to answer multiple-choice items about the scientific method, or delineate the steps involved in the scientific method. Nonetheless, this type of test, however well-written it may be, does not directly assess students' understanding of the scientific method. Understanding implies not only that students can describe important components of the method, but also that they are able to apply the method in a variety of situations. Rather than administering a pencil-and-paper task that measures aspects of the scientific method, the teacher could observe students carrying out an experiment using the scientific method in order to determine their level of understanding. This observation would be a more direct measure of students' understanding of the scientific method. Performance assessments, therefore, include an element of directness that is rarely found in the types of tests typically employed in education measurement (Frederiksen & Collins, 1989; Resnick & Resnick, 1991; Wiggins, 1989).

Performance versus Traditional Tests

Renewed interest in performance assessments stems in large part from dissatisfaction with more traditional types of tests, particularly those comprised of multiple-choice items (Frederiksen, 1984; Frederiksen &Collins, 1989; Resnick & Resnick, 1991; Wolf, Bixby, Glenn, & Gardner, 1991). Before discussing the criticisms of traditional testing practices, it may be helpful to differentiate between the functions that

tests may serve in the educational system. According to Resnick and Resnick (1991), there are three main classes of educational testing: (a) public accountability and program evaluation, (b) instructional management and monitoring, and (c) student selection and certification. Public accountability permits "those in positions of public oversight and responsibility for the education system to monitor the schools' performance....[and] are intended to provide those responsible for the funding and civic supervision of education with information on how the school system as a whole is performing" (p. 48). Program evaluation provides "information on whether a particular instructional or support program is succeeding in its academic goals" (p. 48). Due to their ease of administration and scoring and their breadth of coverage, the tests used for public accountability and program evaluation purposes have typically been standardized achievement tests.

A second function of testing is to select and certify students for particular educational institutions or programs. The Scholastic Aptitude Test (SAT) and the American College Test (ACT) are examples of tests used to select students for college admissions. Both the SAT and ACT are multiple-choice tests that are intended to measure very general learning that is not associated with any particular curriculum content (Resnick & Resnick, 1991, p. 49).

Lastly, tests may be used for the purpose of instructional management and monitoring of student progress, which includes the assigning of students to classes, the diagnosis of students' particular strengths and weaknesses, and the monitoring of the effectiveness of small units of instruction (Resnick & Resnick, 1991). More varied in form

than those used for accountability and selection purposes, these tests are typically designed by the teacher and are more closely aligned with the curriculum.

Standardized, multiple-choice tests used for public accountability and program evaluation purposes are the most criticized form of assessment. These types of tests are criticized for a myriad of reasons, but one of the most compelling arguments against their use is the suggestion that they have an extremely negative effect on curriculum (Frederiksen, 1984; Frederiksen &Collins, 1989; Resnick & Resnick, 1991; Wolf et al., 1991). This negative effect is rooted in the indirect nature of most standardized tests: many researchers fear that in preparation for accountability tests, educators have begun to teach and promote the indicators of ability, rather focusing on the actual abilities of interest (Resnick & Resnick, 1991). This would occur, for example, when an English teacher preparing students for the California Achievement Test (CAT; CTB Macmillan/McGraw-Hill, 1985) replaces class discussion of works of literature with classroom exercises designed to help students identify the correct answer from four alternatives summarizing the theme of a written passage. In this scenario, the teacher is reacting to the concern for accountability by restricting her curriculum to the content area and the type of item found on the standardized multiple-choice tests. Because these test items typically measure indicators of ability, the teacher has inadvertently begun to teach the indicators of skills and competencies rather than the skills and competencies themselves. In order to reverse this deleterious trend, many researchers suggest that traditional multiple-choice tests must be replaced with assessments "that directly reflect and support the development of aptitudes and traits they are supposed to measure"

(Frederiksen & Collins, 1989, p. 28). Using direct measures will lead to instructional focus on the aptitudes of interest rather than on the indicators of those aptitudes.

In addition to their negative impact on curriculum, multiple-choice tests have also been criticized for their emphasis on the assessment of lower-level skills, factual knowledge, memorization of procedures, and isolated skills (Frederiksen & Collins, 1989). Furthermore, many claim that multiple-choice items neither engender nor assess the types of activities that most educators would agree are central to learning and growth, such as complex thinking, thoughtfulness, imagination, pursuit, inference, analysis, and interpretation (Frederiksen, 1984; Wolf et al., 1991). Others are disenchanted with standardized tests because they fail to display students' patterns of thought: these tests are typically administered to students on a single occasion and under a timed condition (Wiggins, 1989). In addition, the arbitrary time constraints imposed during administration result in a test that "privileges speed and efficiency above contemplation and accuracy" (Wolf et al., 1991, p. 45). Standardized tests also have been accused of treating students "as objects - as if their education and thought processes were similar and as if the reasons for their answers were irrelevant. Test-takers are not, therefore, treated as human (emphasis added) subjects whose feedback is essential to the accuracy of the assessment" (Wiggins, 1989, p. 708). Other common criticisms include the lack of feedback provided by standardized test (Wiggins, 1989), the emphasis on a right/wrong scoring dichotomy (Resnick & Resnick, 1991; Wiggins, 1989), intolerance of collaboration (Wiggins, 1989), discouragement of teacher-student dialogue (Wiggins, 1989), and the decontextualized

nature of most standardized tests (Resnick & Resnick, 1991; Wiggins, 1989; Wolf et al., 1991).

Current State of Performance Assessments

In response to these criticisms, performance assessments have been suggested and promoted by many as a viable alternative to traditional, multiple-choice standardized tests, and the development, implementation, and evaluation of performance assessments has become the focus of many journals articles over the past two decades (for example, see special issue of Educational Measurement: Issues and Practice, 1995, Volume 14) . State governments appear to have noticed the attention given to performance assessment: at least 75% of the states now supplement or supplant their traditional tests with some form of a customized performance assessment (Resnick & Resnick, 1991). To date, the most common form of performance assessment implemented on a statewide basis has been the writing assessment (Resnick & Resnick, 1991). Thirty-seven states now administer some form of writing assessment (Council of Chief State School Officers, 1998). Typically, these examinations require students to write essays that are awarded quantitative scores by panels of trained judges. Although the state writing assessments vary in the type of question asked, the length of the essay, scoring criteria, and degree of imposed structure, they are believed by many to be vast improvements over the writing batteries found in most standardized, multiple-choice achievement tests (Resnick & Resnick, 1991). Furthermore, "national and state testing agencies are recognizing that students' open-ended responses can be scored with sufficient reliability to provide data on the quality of learning" (p. 60) and these tests have now shown "the feasibility of using complex,

integrated performances, rather than series of isolated questions, in a public accountability program" (p. 60).

Variables Affecting Student Performance on Tests

Gender and Race

Early studies and evaluations of statewide performance assessments have tended to focus on one of two issues: the logistics of developing, marketing, implementing, and scoring the performance assessments, or the psychometric issues (e.g., reliability, validity, generalizability) related to this (or any) form of educational measurement (e.g., Koretz, Stecher, Klein, & McCaffrey, 1994; LeMahieu, Gitomer, & Eresh, 1995). Thus far, these logistical and psychometric issues have received substantial coverage in the professional literature. What has received significantly less coverage is discussion of how systematic and personal variables affect students' performance on these types of tests. For example, relatively few studies have carefully examined the effect that school variables (e.g., curriculum, teacher attitudes) have on scores of performance assessments.

A few studies have focused on how race and gender interact with scores on performance assessments; for example, research on gender differences in assessed quality of writing indicate that when these differences are evident, girls tend to perform better on writing assessments than do boys (Engelhard, Walker, Gordon, & Gabrielson, 1994; Hyde & Lynn, 1988; US Department of Education, 1996). These differences have been found in very young writers (Nicolopoulou & Scales, 1990, as cited in Engelhard et al., 1997) and in high school writers (Barnes & Barnes, 1990, as cited in Engelhard et al., 1997). The effect of gender on writing performance has also been shown to be mediated by the type

of writing task (e.g., narrative, descriptive, or expository mode of discourse), with the largest difference found in descriptive writing probes (Engelhard et al., 1994). The higher performance of girls has been explained in numerous ways, including research showing that girls tend to use more precise and exact grammar than boys, that girls "are more likely than boys to have mastered the language of the schools as reflected in the statewide assessment of written composition" (Engelhard et al., 1994, p. 206) and that girls have more direct and personal writing styles than boys (Engelhard et al., 1994).

Race/ethnicity is a second student characteristic that has been examined in terms of its effect on performance assessments (Engelhard et al., 1994; US Department of Education, 1996). Studies of statewide and national writing assessment programs demonstrate that, in general, African-American students receive lower ratings than do Caucasian students (Engelhard et al., 1994; US Department of Education, 1996). Like that of gender, the effect of race/ethnicity is sometimes mediated by the type of writing task presented, with the largest difference found in narrative writing tasks (Engelhard et al., 1994). Researchers have suggested numerous possible explanations for these differences, including findings related to the disproportionate number of African-American students in lower-track, non-writing curricula (Irvine, 1990; Oakes, 1985), differences in the way that English teachers interact with African-American and Caucasian students (Irvine, 1990), and differences in language usage between races (Irvine, 1990). In many documented cases, performance assessments (including writing assessments and portfolio assessments) have been less effective than traditional assessments in lessening the score differences between African-American and Caucasian students (Dunbar, 1987; LeMahieu, Gitomer, &

Eresh, 1995; Linn, Baker, & Dunbar, 1991). Further research is needed to explain the difference between African-American and Caucasian students' performance on writing tasks, as well as investigations into the performance of other minority racial groups (e.g., Hispanic, Native American).

Cognitive Ability, Personality, and Style

Although a few studies have examined how student demographic variables such as gender and race impact success on performance assessments, virtually no studies have been conducted on the personal qualities that students bring into the testing situation in terms of how they may affect achievement on performance assessments. Personal qualities include variables such as personality, temperament, interests, learning style, and energy level. Along with cognitive ability, these personal characteristics have also been examined in attempts to gain insight into the factors that impact success of children and adults in the home, school, and work environments.

Of the aforementioned personal qualities, it is undoubtedly cognitive ability that has received the most attention in studies of academic and occupational success (e.g., Neisser et al., 1996). Cognitive ability is certainly predictive of academic success: the vast majority of research on the relationship between scores on tests of cognitive ability and academic achievement suggests that the correlation between the two ranges from approximately .40 to .70, depending upon the measures used (Flanagan, Andrews, & Genshaft, 1997). This estimate shows that intelligence accounts for approximately 15 to 50% of the variance in academic achievement. In addition, ability (as measured by intelligence and aptitude tests) has also been demonstrated to be predictive of job

performance. For example, Herrnstein and Murray (1994) found that over many tests and many jobs, the average correlation between job productivity and ability is .40. This means that 16% of the variance in job productivity is accounted for by variance in ability. The question, then, is in addition to error, what else may account for the remaining 50 to 85% of variability in academic achievement, and the remaining 84% in job performance? Psychologists have turned to two other personal characteristics, personality and style, to help answer this question. To clarify, ability refers to how well someone can do something. Personality (used synonymously here with "trait") has been defined as "a relatively stable disposition to engage in particular acts or ways of thinking" (Kamphaus & Frick, 1996, p. 1). To illustrate, it is not unusual to hear a person described as "an optimistic girl" or a "reclusive boy." This optimism or reclusiveness is conceptualized as being part of the individual's personality, and is generally thought to impact to some degree the individual's proclivities and successes in academic and occupational affairs. The third factor, style, lies somewhere along a continuum between ability and personality. In contrast to ability (how well someone can do something), a style represents a set of preferences. The distinction between style and personality, however, is more obtuse. According to Sternberg (1997b), a specific term, cognitive style, was developed by cognitive psychologists conducting research into problem solving and sensory and perceptual abilities (p. 134). Research on cognitive style was greatly expanded after World War II at Brooklyn College, the Menninger Foundation, and the Fels Institute, where psychologists were investigating the variables the affect information processing, such as perception, organization, and retention of input (Keefe, 1987).

To elucidate the distinction between ability, personality, and style, consider the example of two young men. Both men have math abilities that have been measured to be in the superior range, and both perform equally well in an identical math curriculum. In addition, both men are identified as "introverted" by a widely-used measure of personality. This introversion is considered to be a stable trait of both men, and is not conceptualized as being a choice of either. In terms of careers chosen by both men, one opts to be an accountant while the other chooses to be a mathematician. Why? Certainly both professions require a certain amount of mathematical ability. Neither accounting nor mathematics necessarily requires extroversion to the degree that a field like sales would require. What may contribute, then, to the difference in career choice? It is easy to postulate that a number of variables may have affected this choice: parental wishes, availability of training programs, or financial and social aspirations. It is also possible that the men had different cognitive styles, or preferences for how they approached tasks. For example, one man may prefer to dwell in the world of abstract theories and proofs, while the other prefers tasks that are less ambiguous and more pragmatic. The distinct preference of each man may be conceptualized as his cognitive style. In this example, it is the cognitive style that impacts the different career choices, not the ability or the personality of each man.

In this illustration, cognitive style refers to the degree of abstractness and pragmatism in the types of task that each man prefers. However, there is no definitive theory of cognitive style and what it entails. The definition of cognitive style as being the way a person processes information is vague, and is not shared by all who use the term.

Over the last half-century, at least 20 specific theories of style have been presented, some possessing more empirical backing than others, but all of which attempt to bridge the study of ability and the study of personality. Most of these theories fall within one of three categories: cognition-centered styles (how we perceive), personality-centered styles, or learning styles (Sternberg, 1997b).

Previous Research on Styles

Cognition-Centered Styles

Field dependence-independence. The cognition-centered styles movement began in the 1950s and early 1960s, and included, among others, the theories of field dependence-independence, equivalence range, category width, conceptual style, and impulsivity-reflectivity (Keefe, 1987; Sternberg, 1997b) . Field dependence-independence theory emerged from studies of how perceptual function relates to personality. Many projective tests, such as the Rorschach (Rorschach, 1921), are essentially measures of perception, or of how a person perceives external stimuli. While studying perception, Witkin and Goodenough (1981) recognized that people differed in their ability to locate an upright and to recognize stimuli embedded in conflicting contextual clues. Witkin developed two tests to measure these abilities, the Rod and Frame Test (RFT) and the Embedded Figures Test (EFT). The RFT requires the individual to ignore a visual context (the frame) in order to locate a true vertical rod. In the EFT, the individual must recognize a previously seen shape within a larger design that obscures or embeds the original. The field-independent person can recognize the true vertical regardless of the tilt of the frame, and can identify the original figure embedded in the more complex design. This occurs

presumably because the field-independent person does not require contextual clues the identify stimuli. Witkin found that, in contrast, the field-dependent person would tend to orient the rod in relation to the tilted frame on the RFT and to have more difficulty recognizing shapes embedded in complex designs on the EFT. Witkin noted that an individual's way of perceiving tended to be stable across many perceptual tasks, an observation which led him to study the personal characteristics that may be correlated with field-dependence and independence. As he moved beyond the study of visual perception, Witkin and others began to examine how style was related to a number of personality variables such as leadership (Weissenberg & Guenfeld, 1966) and social conformity (Witkin, Dyk, Faterson, Goodenough, & Karp, 1974). Other research focused on the relationship between field dependence-independence and interpersonal relationships (Witkin & Goodenough, 1977), learning and memory (Goodenough, 1976), math achievement (Vaidya & Chansky, 1980), specialization in college and graduate school (Raskin, 1985), and cross-cultural differences (Berry, 1976). As the breadth of the studies increased, the name of the theory was changed from field dependence-independence to global versus articulated cognitive style. In a review of the literature on this theory, Anastasi (1988) concluded that:

> field-independent persons tend to follow active, "participant" approaches to learning, while field-dependent persons more often use "spectator" approaches. In interpersonal situations, on the other hand, the field-dependent tend to have certain advantages in getting along with others. They tend to be more attentive to social cues, more responsive to other persons' behavior, and more emotionally open than are the field-independent persons. It appears that neither end of the field-dependent-independent continuum is necessarily or uniformly favorable or unfavorable; rather, the value of deviations in either direction depends on the demands of particular situations. (p. 627)

Studies examining the relationship between field dependence and intelligence have reported correlations in the .40 - .60 range (Goldstein & Blackman, 1978). In particular, performance on measures of field dependence have been correlated with scores on the Wechsler Intelligence Scale for Children (WISC) subtests of Block Design, Picture Completion, and Object Assembly (Goodenough & Karp, 1961). These findings suggest that field dependence-independence may be more of an ability than a true style, or at the very least, that there is a relationship between field dependence-independence and ability.

Equivalence range. In the early years of research into cognitive styles, Witkin's theory of field dependence-independence received more attention than any other (Witkin & Goodenough, 1981). However, other theories falling within the category of cognition-centered styles also received attention in the 1950s and early 1960s. These include the lesser-known theories of equivalence range, leveling-sharpening, category width, and impulsivity-reflectivity. The theory of equivalence range was originated by Gardner, Jackson, and Messick (1960), who defined it as "the degree of differentiation in individuals' experiencing of similarity and difference" (Goldstein & Blackman, 1978, p. 6). Gardner (1953) hypothesized that individuals would be consistent in "what they will accept as similar or identical in a variety of adaptive tasks" (as cited by Goldstein & Blackman, 1978, p. 6). The concept of equivalence range was later extended to that of conceptual differentiation, which is the term currently used (Gardner & Moriarity, 1968). Conceptual differentiation refers to the tendency to either see things that are very much the same as different, or the tendency to see things that are very different as the same. Conceptual differentiation was measured by the Object Sorting Test, a task that requires

subjects to sort 73 objects into as many categories as the subject believes are warranted (Goldstein & Blackman, 1978).

Leveling versus sharpening. Leveling versus sharpening is a cognitive style that describes individual differences in memory processing (Gardner, Jackson, & Messick, 1960; Keefe, 1987). People who assimilate new events into events that have already been perceived and stored in memory are called "levelers." Those demonstrating less assimilation are called "sharpeners." In other words:

> levelers tend to blur similar memories and to merge new experiences with previous ones; they tend to over-generalize. Sharpeners are able to distinguish small differences and to separate memory of prior experiences more easily from current ones; they tend to over-discriminate. (Keefe, 1987, p. 9)

Measures used to assess leveling-sharpening include a test that requires subjects to judge the sizes of squares of light of increasing size. In general, subjects tend to underestimate the sizes of the squares as new, larger squares are presented. Levelers tend to make larger underestimates because of the more salient trace aggregate of previously-seen squares (Gardner et al., 1960; Goldstein & Blackman, 1978).

Category width. Category width refers to the tendencies people make when estimating the range of amounts, widths, lengths, and other units of measurement (Pettigrew, 1958). According to this theory, when giving estimates of ranges people have a tendency to give either broad or narrow ranges. Broad categorizers prefer to include many items and lessen the risk of leaving something out, while narrow categorizers prefer to exclude doubtful items and lessen the probability of including something that does not fit (Keefe, 1987, p. 9). To measure category width, Pettigrew developed the 20-item C-W

test. The C-W test presents the subject with a central measure and then requires selection of one of four alternatives to represent the highest end of the range and one of four alternatives to represent the lower end of the range (Keefe, 987; Pettigrew, 1958). Explanations for the observed consistency in subjects' judgments of category width are not conclusive (Goldstein & Blackman, 1978). Nor has research extended to measure category width when it applies to estimates of ranges of psychological variations (Sternberg, 1997b).

Reflectivity-impulsivity. The concept of reflectivity-impulsivity was originally introduced by Kagan, Rosman, Day, Albert, and Phillips (1964) to describe "the differences in the speed in which [children] make decisions under conditions of uncertainty" (Goldstein & Blackman, 1978, p. 11). "Impulsive" children tend to give the first answer they think of, allowing themselves to make mistakes. In contrast, "reflective" children tend to consider alternate solutions before deciding, and are more careful about making mistakes (Keefe, 1987). The Matching Familiar Figures Test (MFFT; Kagan et al., 1964) was developed to measure reflection-impulsivity by presenting the subject with 12 test stimulus pictures and asking the subject to choose (among several alternatives) the picture that exactly matches the stimulus picture. A child who responds rapidly and with more errors is considered impulsive, while the child who responds slowly and with less errors is labeled reflective. Most of the studies with the MFFT report a negative correlation between response rate and errors, with a median correlation across studies of approximately -.48 (Messer, 1976). Scores on the MFFT have been found to be moderately related to scores on Witkin's EFT, with reflective individuals demonstrating

more field independence than the impulsive type (Messer, 1976). MFFT error rates are also negatively correlated with age, suggesting that reflectivity increases with age (Messer, 1976). Lastly, a review of approximately 23 studies found a median correlation of .16 between MFFT response-time and scores on eight widely-used test of intellence (Messer, 1976). The correlation between errors on the MFFT and intelligence test scores has been estimated as approximately -.30 (Messer, 1976). These results suggest that there is a moderate relationship between decision-making speed and intelligence, with more reflective children scoring higher on intelligence tests. These results, however, are confounded with order of test presentation and the type of test used to measure intelligence (e.g., verbal, nonverbal).

Summary of theories of cognition-centered styles. Other cognition-centered styles emerged during this time, but the theories of field dependence-independence, equivalence range, leveling-sharpening, category width, and impulsivity-reflectivity are probably the most well-known to emerge during this era. Cognition-based theories that received less attention but are worth mentioning include: scanning (Gardner & Moriarity, 1968), which is the extent to which an individual attempts to verify the judgments he makes (Goldstein & Blackman, 1978, p. 9); conceptual style, which is "stable individual preferences in mode of perceptual organization and conceptual categorization of the external environment" (Kagan, Moss, & Sigel, 1963, as cited in Goldstein & Blackman, 1978, p. 10); and cognitive complexity versus simplicity, which contrasts persons who accept diversity and conflict in a multi-dimensional world versus those who prefer consistency and regularity in the environment (Bieri, 1961; Scott, 1962).

The cognitive-centered theories emerging in the 1950s and 1960s marked the beginning of research on styles in the field of psychology. In general, these initial studies of style investigated how the mind actually processes information and how the mind is affected by the perceptions of the individual. Researchers in this area assumed that the way that an individual processes sensory stimuli affects the way in which that person interacts with the world. This assumption is very congruent with an information-processing theory of thinking, which draws the analogy between a computer and the human mind. Prior to the emergence of these earliest cognitive-style theories, psychologists and physiologists had been speculating on the relationship between mind and body, and most accepted that there was, in fact, an interaction between the two. As research on this interaction evolved, the logical outgrowth was an interest in discovering discrete "types" of how input is processed, and the subsequent effects that these types have on overt human behavior. Many of the early instruments designed to measure cognitive style reflected this emphasis on perception by including items that relied on visual sensory input (e.g., the EFT). As research on styles continued, the instruments developed to measure them came to reflect the de-emphasis on sensory perception in lieu of psychological constructs that were less related to or controlled by sensory input. The psychology-centered and (action-centered) styles discussed next exemplify this shift in research.

The theories and instruments related to cognition-centered styles were not directly concerned with education in terms of how styles may affect teaching and learning. That type of research emerged over the following decades when researchers began examining the impact of personality and learning styles on educational variables such as type of

instruction delivered and preferred modes of learning. The growth of new theories of style following the early 1960s were accompanied by increasingly sophisticated methods of statistical analyses and higher standards in research methodology, leading to a more rigorous examination of the psychometric properties of the newer instruments and the utility of the theories.

Personality-Centered Styles

As mentioned before, styles may be visualized as falling somewhere on a continuum between ability (how well someone can do something) and personality (a disposition to engage in a certain way if thinking).In contrast to "cognition-centered" theories, which emphasized how perception influences thinking, personality-centered styles are more concerned with how personality traits affect thinking and behavior. Measures of personality-centered style are often distinguished in the literature from actual "personality tests" (e.g., Sewall, 1986; Sternberg, 197b), although they are overlapping concepts.

The Myers Briggs Type Indicator (MBTI; Myers & McCaulley, 1985) is undoubtedly the most well-researched instrument of personality type in existence. A recent review of the literature produced over 285 references to the MBTI, and one source reported that over 2 million copies of the MBTI are now sold annually (Pittenger, 1993). The MBTI was developed over a period of 35 years by Isabel Briggs Myers and her mother Katharine C. Briggs, after Briggs became interested in personality theory as a means to analyze characters in literature. Briggs became increasingly interested in Jung's (1921/1926) theory of psychological type and believed that his theory could provide a link

between personality and job performance. During World War II, Myers and Briggs noticed that many people were taking jobs that were mismatched with their perceived types. They decided to develop a test for personnel selection that was not based on aptitude. Beginning in 1942, Myers developed several forms of the MBTI and proceeded to test items on over 15,000 high school and college students. In 1962 the MBTI was published by Educational Testing Service as a research instrument. By the late 1960s and 1970s the MBTI was becoming extremely popular with counselors and psychologists. In 1975, the MBTI was published by Consulting Psychologists Press as a proprietary instrument. Today the MBTI is used in a wide variety of settings and for a wide variety of purposes, including individual, group, and family counseling, academic and career counseling, education (e.g., to identify learning styles, teaching styles, aptitude, achievement, and motivation), and organization (leadership training, career planning, teamwork, participative management).

Myers developed the MBTI to apply Jung's theory of psychological type to an analysis of nonclinical populations. Jung's theory emerged from his clinical work and he explained its purpose as follows:

> If one is plunged, as I am for professional reasons, into the chaos of psychological opinions, prejudices, and susceptibilities, one gets a profound, indelible impression of the diversity of individual psychic dispositions, tendencies, and convictions, while on the other hand one increasingly feels the need for some kind of order among the chaotic multiplicities of points of view. This need calls for a critical orientation and for general principles and criteria, not too specific in their formulation, which may serve as "points of departure" in sorting out empirical matter. (1971, p. xiv)

The MBTI is an attempt to categorize or "type" people according to their preference in

each of four functions: attitudinal, judgmental, perceptual, and interpretive. Each of these functions has two mutually exclusive components, one of which is the dominant feature of the individual's personality (Pittenger, 1993). The attitudinal component describes how we interact with others, and includes the dichotomy of extroversion/introversion. Extroversion characterizes those with an interest in people and the environment, and those who give weights to events in the external world. Introversion describes people who engage more in internal activities and use concepts and ideas to understand events. The second function, perception, separates intuitive and sensing people. A person with an intuitive preference relies on unconscious perceptual processes, and is interested in "future possibilities, implicit meanings, and symbolic or theoretical patterns suggested by insight" (McCaulley, 1990, p. 183). In contrast, a sensing person relies more on direct perception and that which is considered to be real, immediate, and practical. The judgment component of personality can be either thinking or feeling. Thinking types use logical and rational processes to reach decisions, whereas feeling people tend to be more subjective in their decision-making and to consider the more affective consequences to events. The fourth distinction refers to how people interpret information. Those with a perceptive preference "enjoy being curious and open to changes, preferring to keep options open in case something better turns up" (McCaulley, 1990, p. 183). Judging people, on the other hand, tend to be decisive and enjoy "organizing, planning, and structuring" (McCaulley, 1990, p. 183).

The MBTI standard instrument consists of forced-choice questions that inquire about individual's preferences in certain behavioral situations. The instrument was

developed for individuals sixth grade and older, and can be administered individually and to groups. An indicator for children in Grades 2 through 8 was published in 1988. The final personality profile derived from the scores consists of the four letters that represent the preferences for each type dimension. For example, a person with an INTJ type would be introverted, sensing, thinking, and judging. There are a total of 16 potential personality types, each considered to be qualitatively unique and to represent specific cognitive and affective clusters.

Over the years the MBTI has been subject to numerous studies of its psychometric properties (see Pittenger, 1993; Sewall, 1986), producing somewhat equivocal findings. Of particular interest are the studies comparing MBTI types to measures of academic performance. Rollins (1990) predicted that sensing students would prefer more action-oriented course work (e.g., laboratory assignments, projects), whereas students of the intuitive type would prefer more passive assignments (e.g., reading, working independently). Contrary to the hypothesis, Rollins found that intuitive and sensing students enrolled in a high school agricultural course gave nearly identical rankings for 19 different learning activities.

A study of the match between teachers' and students' MBTI types found that for course evaluations and instructor evaluations, the lowest ratings were given when a type mismatch existed (Cooper & Miller, 1991). In the same study, no significant relationship was found between the students' grade and the match between personality type.

In terms of predictive validity, the MBTI manual is replete with correlational studies focused on field of study, occupation, and subspecialty. Most of this research was

conducted with adults and is most relevant for vocational and career counseling purposes. Significantly less research has examined how MBTI type relates to academic performance in school-age children. Thorne and Gough (1991) administered the MBTI to 99 male and 99 female Berkeley sophomores and correlated their types with the College General Vocabulary Test (Gough & Sampson, 1954), the Verbal and Mathematical scales of the Scholastic Aptitude Test (SAT), and students' self-report of high school and college grade point average (GPA). The authors found significant correlations between the following: males' scores on the General Vocabulary Test and the Extroversion/Introversion (EI) dimension of the MBTI ($\underline{r}$ = 34), and females' scores on the General Vocabulary Test and the EI ($\underline{r}$ = .21) and Sensory/Intuitive (SN) dimensions ($\underline{r}$ = .28). For the EI and SN dimensions, those with high Introversion (versus Extroversion) and Intuitive (versus Sensing) scores tended to have higher vocabulary scores. The verbal component of the SAT and the EI and SN dimensions also correlated. Higher verbal SAT scores for males were related to higher Introversion scores ($\underline{r}$ = .46), and higher verbal SAT scores for females were correlated with Intuitive scores ($\underline{r}$ =.25). In terms of self-reported GPA, a significant negative correlation was found between high school GPA and the Judging/Perceiving (JP) dimension of the MBTI for males ($\underline{r}$ = .26), and a significant positive relationship was found between the college GPA of females and the SN dimension ($\underline{r}$ = .28). These results suggest that males with high scores on Perceiving (versus Judging) tended to report lower high school GPA's, and that females with high Intuitive scores tended to have high college GPA's.

Thorne and Gough's study does not provide strong evidence for how (or if) the MBTI is related to measures of academic performance, however. Many sources are available on how MBTI type may affect students' performance in the classroom (for example, see Mamchur, 1996 for a teacher's guide on how type presumably affects instruction), but very little empirical research exists to substantiate the claims. The MBTI appears to have substantial face validity, as evidenced by its widespread use, but significantly less psychometric support.

Learning Styles

Introduction to learning styles. None of the cognitive-centered or personality-centered theories of style discussed thus far were developed specifically to answer questions related to school learning. Although several instruments based on these theories have been applied in studies examining academic variables (e.g., the MBTI), these studies are few in number and have had minimal impact on instructional practices. Since the 1970s, however, several theories and models of "learning style" have emerged that focus specifically on variables that impact students' school success. In 1979, the National Association of Secondary School Principals (NASSP) published a guide for diagnosing learning styles and prescribing appropriate instructional strategies. In this guide, NASSP defined learning styles as "characteristic cognitive, affective, and physiological behaviors that serve as relatively stable indicators of how learners perceive, interact with, and respond to the learning environment"(as cited in O'Keefe, 1987, p. 5). Dunn and Griggs (1988) defined learning style as "a biologically and developmentally imposed set of characteristics that make the same teaching method wonderful for some and terrible for

others" (p. 3). They hypothesized that this set of characteristics is unique to each student and dramatically impacts the probability that a student will learn given a particular learning milieu. In fact, proponents of learning style assert that, if students' preferred modes of learning are accommodated in the classroom, "almost any student can learn or master the material set before them" (Henson & Borthwick, 1984, p. 4) regardless of the students' aptitudes. This assertion contrasts sharply with the view that intelligence places heavy limitations on levels of learning, a view widely held since E.L. Thorndike's 19th century studies on learning (Henson & Borthwick, 1984).

At least a dozen measures of learning styles have been published in the last several decades, some of which have received more attention than others. The following three theory-based measures of learning style have received considerable attention in the literature and warrant further description: the Gregorc Style Delineator (GSD; Gregorc, 1982), the Kolb Learning Style Inventory (KLSI; Kolb, 1978), and Dunn and Dunn's Learning Style Inventory (LSI; Dunn, Dunn, & Price, 1979).

Gregorc style delineator. The GSD is a self-administered and self-scored measure developed for use by adolescents and adults for "self-analysis, for self-observation, and for prompting understanding of self, others, and environments" (Gregorc, 1984, p. 29). Gregorc asserted that learning style consists of "distinctive, observable behaviors that provide clues to the functioning of people's minds and how they relate to the world" (Sewall, 1986, p. 10). The instrument is designed to identify which of four learning styles an examinee is most aligned with: Concrete Sequential, Concrete Random, Abstract Sequential, and Abstract Random. The Concrete/Abstract dichotomy refers to "reference

points for thinking" or "space," whereas Sequential/Random is a temporal or ordering component (Gregorc & Ward, 1977, p. 21; Sternberg, 1997b). Although the Gregorc Style Delineator is frequently cited in the literature, its psychometric properties are unknown. In fact, the test manual provides no information on the development of its norms, which renders interpretation of the raw scores problematic. In addition, the one stability study that is described is flawed in terms of unequal test-retest intervals across subjects (Sewall, 1986). Evidence for validity of the instrument is also tenuous, and allows for little confidence in what the test is actually measuring. For example, there is no statistical support for the four constructs of the instrument, nor is there evidence of expert testimony or a psychological theory that could potentially add support to Gregorc's "styles." In addition, the studies provided as evidence of the predictive validity of the GSD used criterion measures (e.g., a questionnaire developed by Gregorc) that have unsubstantiated validity themselves. A review of the literature reveals a paucity of empirical evidence for this measure. One recent reviewer of the GSD concluded:

> ...the Gregorc Style Delineator appears to have little practical value to the individual seeking a better understanding of their personal learning style. I believe that the most appropriate use of this instrument would be to provide an example of how not to construct a assessment tool. (Sewall, 1986, p. 57)

Kolb's learning style inventory. Like the GSD, Kolb's Learning Style Inventory (KLSI) is also self-administered and self-scored, and is intended for adolescent and adult readers. According to Kolb's experiential learning theory (upon which the LSI is based), there are four learning styles: Converging, Diverging, Assimilating, and Accomodating. Convergers tend to use deductive thinking in conceptualizing abstract matters, and prefer

to focus on specific problems. In contrast, Divergers prefer concrete experiences and approach their dealings with others in a more imaginative and emotional fashion. Assimilators are reflective observers, like to create theoretical models, and tend to use deductive reasoning to assimilate differing pieces of information. Accomodators like concrete experiences and active experimentation, and enjoy taking risks (Sternberg, 1997b, p. 145).

Kolb's model and instrument were designed for adult organizational systems and management training. Most of the studies on this instrument also have been conducted within these settings (DeBello, 1989). A recent review of the LSI reveals several weaknesses, including limited evidence regarding the test's reliability and validity (Sewall, 1986). The LSI was normed primarily on a group of management college students. The methodology used and descriptions of the normative sample are described as "very limited," and the construct validity information presented in the test manual is called "speculative" (Sewall, 1986, p. 24). No significant, consistent relationships have been found between scores on the LSI and other measures of learning styles (Sewall, 1986). In terms of predictive validity, studies of the value of the LSI in predicting specific career choices within the medical field have yielded equivocal results. Studies of the relationship between the LSI and specific instructional strategies have largely been anecdotal and inconsistent in their findings. The lack of evidence supporting the validity and reliability of the LSI limits its usefulness for any purpose other than informal self-analysis.

Dunn and Dunn learning style inventory. Dunn and Dunn identified the 18 elements that they believe constitute learning style, and upon which their instrument, also

called the Learning Style Inventory (LSI), is based (Henson & Borthwick, 1984; Keefe, 1987). The 18 elements are further organized into four major elements: environmental, emotional, sociological, and physical. Environmental elements include sound, light, temperature, and design. Emotional elements include motivation, persistence, responsibility, and structure. The sociological factor include preferences for learning alone, in a pair, with peers, as part of a team, and with an authoritative figures. Lastly, physical elements include perceptual modalities, such as daily periods of energy and an individual's need for physical mobility. According to the Dunns' theory, these elements affect each learner to a different degree: those elements that strongly affect an individual are called "strong preferences," and those that are less influential for a particular learner, simply "preferences."

The Dunns' LSI is a 100-item, self-administered questionnaire designed for students in Grades 3 through 12. The inventory was developed to diagnose learning styles so that the most appropriate teaching strategy may be prescribed. According to theory, students have the optimal opportunity of learning when their learning style is accommodated in the classroom. At the current time, the Dunns' LSI is the most widely researched learning styles instrument, and it reportedly has the most substantiating evidence of reliability and validity of all instruments of its type (DeBello, 1989). The Ohio State University's National Center for Research in Vocational Education published the results of a two-year study of learning style instruments and concluded that the Dunns' LSI had "impressive reliability, and face and construct validity" (Kirby, 1979). In his book on learning styles, Keefe (1987) stated that "the [LSI] is a practitioner-oriented instrument

with good reliability and widespread application, particularly in elementary and middle level schools....[but] several of the scales could benefit from better face and construct validity" (p. 24).

The Dunns and other researchers have reported over 40 studies on the utility of their instrument in identifying and prescribing instructional strategies to optimize students' learning. Most relevant for the current study, they have conducted several studies in which students identified as having strong positive or negative preferences were placed in academic situations where they were taught in ways that matched or mismatched their preferences. Dunn concludes that in:

> ...every case, students who were matched with methods, resources, or environments that complemented their reported strong preferences achieved statistically higher; they achieved statistically less well when they were mismatched with their preferences....it is only reasonable to conclude that since students achieve better when taught through their preferences, their preferences must be their strength. (Dunn, 1984, p. 13)

A meta-analysis of 36 studies examining the effect of teaching students through their learning-style preferences as measured by the Dunn's LSI concluded that "students whose learning style are accommodated would be expected to achieve 75% of a standard deviation higher than students who have not had their learning styles accommodated" (Dunn, Griggs, Olson, Beasley, & Gorman, 1995, p. 353). These findings are in direct contrast to an earlier meta-analysis that found that neither modality assessment nor modality instruction are efficacious (Kavale & Forness, 1987). Dunn et al. (1995) criticized the Kavale and Forness meta-analysis for including data from only one study (a master's thesis) on the Dunn's LSI, and also for treating all of the learning style models as

homogeneous. They argue that their meta-analysis provides clear evidence on the utility of using the Dunn's LSI to assess and match students' learning style with the proper type of teacher instruction (Dunn et al., 1995). In 1984, in a special issue of the journal Theory Into Practice devoted to matching teaching to learning styles, editor Kenneth Henson summarizes the crux of the learning style issue with the following statement:

> Matching teaching styles with learning styles is a topic which can spark heated controversy among individuals in any education debate. Some label the movement a shallow excuse for scholarship; others perceive the study as an example of true inquiry, a phenomenological pursuit which goes far beyond analyzing and describing over behavior. (p. i)

Summary of learning style research. In spite of the enthusiasm and dedication with which some researchers have advocated for the study of learning styles, very few studies have earnestly assessed teaching through modality strengths (Dunn, 1984). There are four major reasons for this paucity of attention. First, psychometrically sound measures of learning style are scarce, with the possible exception of the Dunns' LSI. Measures of learning style seem to either completely lack evidence of validity, or the reported evidence is so weak as to be useless. Second, despite the widespread belief that children do, in fact, learn differently, and that these differences should be accommodated in the classroom to maximize success, there is still insufficient evidence regarding the relationship between learning styles and other educationally-related factors (e.g., test performance, students' motivation). This may stem from the first problem, the lack of sound instruments to measure learning style. Certainly, an unreliable measure or one that is not truly tapping into the construct of interest (i.e., learning style) is less likely to demonstrate a significant link to the outcome measures. Another possible explanation could be that the measures of

outcome variables are unsound. For example, in research examining the relationship between students' learning style and motivation, the instrument used to measure motivation is of tenuous reliability and validity. This psychometric weakness lessens the probability that a significant relationship will be found. Third, it is difficult to conduct controlled experimental studies examining the interaction of learning styles and instructional strategies, because most of the measures and theories of learning styles are not prescriptive -- that is, they do not detail the type of instruction that maximizes the performance of students with a particular style. In addition, it is difficult to conduct a study where a teacher uses a pure form of some particular "matched" instructional strategy. For example, if one were testing the veracity of Kolb's accomodating/assimilating dichotomy, how would an educator teach a class of identified "accomodators" in a controlled study? How different would this instruction be from one that is aimed toward "assimilators"? Instructional strategy appears to be an extremely difficult variable to control.

Finally, there is a trend in the learning styles research that pervades the majority of studies examining factors that impact students' academic success: that is, the literature has typically used a very narrow definition of "success." What often happens is that the researcher is interested in how a particular variable (e.g., motivation) affects academic success, and spends an inordinate amount of time in the development of a clear conceptualization of "motivation" and a psychometrically-strong instrument to measure motivation. What is typically overlooked in this equation is how "success" should best be measured. Certainly in the large scale studies, measures of success tend to be constricted

to particular forms of academic tests - - multiple-choice, indirect measures of some knowledge or skill. The reasons for this are clear, in that these types of tests (e.g., the CAT have the desirable qualities of reliability, validity, and ease of administration and scoring. The same trend can be found in studies of learning style: a researcher may find that students with a particular style score higher on the outcome measure, which again is usually a score from a standardized test.

In addition to these criticisms, what has largely been ignored in the literature is the possible interaction between students' personal variables (e.g., style) and their performance on various types of tests. And this is especially true for performance assessments due to their relatively recent entry into the measurement field. Recently, Robert Sternberg (1997b) has attempted to address this issue by suggesting that more attention be paid to how what he calls "cognitive style" affects performance on different types of tests.

Sternberg's Cognitive Styles Theory

Students with high intelligence or skill mastery would be expected to score higher on performance assessments than those students with lower intelligence or nominal mastery- - after all, most assessments are designed to reflect the varying levels of student competency. But what else might these assessments be reflecting? Sternberg (1997b) suggested that school achievement may reflect not only a student's ability, but also the student's preferred way of using his/her abilities. He calls this preference a "cognitive style." Sternberg defines cognitive style as:

> ...a way of thinking. It is not an ability, but rather, a preferred way of using the abilities one has. The distinction between style and ability is a crucial one. An ability refers to how well someone can do something. A style refers to how someone likes to do something. (1997b, p. 8)

On the impact that cognitive style may have on achievement, Sternberg (1990) contends that "styles of thinking and learning are every bit as important as levels of ability and that we ignore students' thinking styles at our peril - - and theirs" (p. 367), and that "the teaching methods, the test formats [emphasis added], and the assignments cater to different styles" (p. 367).

Theory of Mental Self-Government

Sternberg's theory of cognitive styles, which he refers to as the "theory of mental self-government,"(1997b, p. 19) uses government as a metaphor for guiding his explanation of the structure, development, and utility of cognitive styles. The basic assumption to the theory is "that the kinds of governments we have in the world are not merely arbitrary and perhaps random constructions, but rather in a certain sense are mirrors of the mind" (Sternberg, 1997b, p. 148). Thus to understand cognitive styles, one needs consider the following five aspects of government: its functions, forms, levels, scope, and leaning.

Function. According to Sternberg (1997b), the mind's functions are analogous to the three branches of the federal government: the legislative, executive, and judicial. Like the legislative branch of the government, the Legislative function of the mind is concerned with creating, formulating, imagining, and planning (Sternberg & Grigorenko, 1993). Thus Legislative people "like to come up with their own ways of doing things...create their own

rules, and prefer problems that are not prestructured or prefabricated" (Sternberg, 1997b, p. 20). Some preferred activities of Legislative people include writing creative papers, designing innovative projects, creating original artworks, and inventing new things. The Executive function of the mind is concerned with implementing and with doing. Executive people like to follow rules, prefer problems that are prestructured or prefabricated, and like to fill in gaps within existing structures rather than creating structures themselves (Sternberg, 1997b; Sternberg & Grigorenko, 1993). Executive people also prefer activities like solving problems, building from designs, writing papers on assigned topics, and learning assigned information (Sternberg, 1994a). The third function of the mind, the Judicial function, is involved in judging, evaluating, and comparing. Judicial people "like to evaluate rules and procedures, and prefer problems in which one analyzes and evaluates existing things and ideas" (Sternberg, 1997b, p. 21), and may prefer activities such as critiquing the work of others, writing critical essays, and giving feedback (Sternberg, 1994a). It is important to note that all people are served by all three functions of mental self-government; however, in each person, one particular function tends to dominate (Sternberg & Grigorenko, 1993).

Form. The forms of government delineate four ways that people approach the world and its problems: through either a Monarchic, Hierarchic, Oligarchic, or Anarchic style. Monarchic people like to focus on a single task until it is completed and will devote all of their attention and energy to that task while ignoring other projects. Hierarchic people like to do several things at once, and will prioritize the time and energy needed for each activity. Those with an Oligarchic style like to do many things at once, and often

have difficulty prioritizing. And finally, the Anarchic style characterizes individuals who take a random approach to solving problems, and who do not like being restricted by systems or rules (Sternberg, 1994a, 1997a).

Level. Mental self-government is also divided into two levels: Global and Local. Global individuals "prefer to deal with relatively large and abstract issues. They ignore or don't like details, and prefer to see the forest rather than the trees" (Sternberg, 1997b, p. 24). The Global person likes working with concepts and ideas, and would probably enjoy an activity such as writing an essay on the meaning of a work of art. Local individuals, on the other hand, are details people-they prefer to work with specifics and concrete examples, and tend to be oriented toward the pragmatics of a situation. An essay describing the details of a work of art and the interactions between those details would appeal to a local person (Sternberg, 1994a, 1997).

Scope. Just like with real governments, mental self-governments need to deal with issues that are both Internal and External in scope. Individuals with an Internal scope like to work alone, and tend to be more introverted, task-oriented, and aloof than those with an External style. Externalists tend to be outgoing, extroverted, and people-oriented, and they prefer to be around others.

Leaning. And finally, leaning refers to whether a person prefers to take a Liberal or Conservative approach to problem-solving. Liberal style characterizes individuals who like to "go beyond existing rules and procedures and who allow substantial change from the way things are currently done" (Sternberg, 1997a, p. 707). In contrast, Conservative individuals like to do things in familiar, traditional ways.

Stability of Cognitive Styles

According to Sternberg, these 13 cognitive styles are not innate, predetermined, or fixed within individuals. Rather, whenever possible, people choose the styles of managing themselves that are most comfortable for them. Thus, instead of possessing only one style, people instead have profiles of styles that may vary somewhat as a function of the immediate environment or the task at hand. These profiles of styles may also vary dramatically across the lifespan, due to both changes in the environment and changes within the person. Sternberg (1997b) states that "the way you think now may not be the way you will think in 10 or even 5 years, and is probably not the way you thought 10 or even 5 years ago" (p. 89).

Cognitive Styles and School Success

Presenting Sternberg's theory within the framework of educational research, the question then becomes one of how a student's cognitive style can affect success in school. Sternberg (1997b) makes two assertions that underscore the potential importance of understanding how cognitive style relates to school success: (a) "schools and other institutions, from households to businesses to cultures, value certain ways of thinking more than others," (p. 8) and (2) "people whose way of thinking do not match those valued by the institutions are usually penalized" (p. 8). Furthermore, he states that "as a society, we repeatedly confuse styles with abilities, resulting in individual differences that are really due to styles being viewed as due to abilities" (p. 12) and that "what is seen as stupidity or intransigence may actually be nothing more than a mismatch between the style of one individual and the style of another" (p. 12).

What are the implications of cognitive style for academic testing? Sternberg (1994b) contends that "our approaches to testing inadvertently reward students who show certain profiles of thinking styles, while punishing students who show other profiles" (p. 170). In other words, the type of test used to assess student achievement may actually be biased against students with cognitive style profiles that are not congruent with that particular test type. For example, the most widely-used form of assessment, the standardized multiple-choice test, is said to "strongly favor test-takers with an Executive, Local style, and favor somewhat those with Internal and Conservative styles as well" (Sternberg, 1994b, p. 178), because the testing situation is a typically a highly structured one where the parameters for solving the problem are explicitly given (Executive), problems are typically narrow in scope (Local), collaboration is discouraged (Internal), and the scoring key favors the conventional wisdom for solving the problems (Conservative). Sternberg extends this argument with an analysis of the various methods of assessment that are typically found in the schools and the cognitive styles that each would tend to favor, and summarizes his argument with the following statement:

> Certain forms of testing benefit students with certain style profiles, whereas other forms of testing benefit students with other style profiles. Therefore, testing should be varied so as to avoid bias. Ability and achievement tests should measure what they are supposed to measure, and not measure those confounded with styles. At the very least, these styles should be measured independently of abilities and achievements. (Sternberg, 1994b, p. 186)

In order to investigate cognitive styles in the classroom, Sternberg and his colleagues have conducted several studies that focus on students, teachers, and the interaction between students and teachers. Four instruments were used to assess cognitive

style in these studies. The Thinking Styles Inventory (TSI; Sternberg & Wagner, 1991) is a self-report measure in which students rate themselves on a number of preferences (see detailed description in Methods section). The Thinking Styles Questionnaire for Teachers (TSQT; Grigorenko & Sternberg, 1993a) measures teachers' preferences for thinking styles in students. An example of an item (rated on a 1-9 scale) taken from this questionnaire is "I want my students to develop their own ways of solving problems" (legislative). The Set of Thinking Styles Tasks for Students (STSTS; Grigorenko & Sternberg, 1993a) measures students' preferences for styles when performing specific tasks. An example of an item from this task is "When I am studying literature, I prefer to (a) follow the teacher's advice and interpretations of author's positions, and to use the teacher's way of analyzing literature" (Executive), "(b) to make up my own story with my own characters and my own plot" (Legislative), "(c) to evaluate the author's style, to criticize the author's ideas, and to evaluate characters' actions" (Judicial), or "(d) to do something else (please indicate in the space below)." The fourth instrument, the Students' Thinking Styles Evaluated by Teachers (STSET; Grigorenko & Sternberg, 1993b), asks teachers to evaluate the cognitive styles of their students. For example, one item from this test is "s/he prefers to solve problems in her or his own way" (Legislative).

The first study examining cognitive styles in the classroom investigated whether teachers' styles differed as a function of the schools (Sternberg & Grigorenko, 1995). The participants were 85 teachers in four diverse schools: a large urban public school, a prestigious secular private school, a Catholic parochial school, and an avant-garde private school. Several findings of this study are noteworthy. First, teachers at lower grade levels

were found to be more Legislative and less Executive than were teachers at higher grade levels. Second, older teachers were more Executive, Local, and Conservative than younger teachers. Third, teachers of different subject matter displayed some interesting differences in cognitive styles - - for example, they found that science teachers tended to be more Local, whereas humanities teachers tended to be more Liberal. Fourth, and finally, when the ideologies of the four schools were rated by an independent researcher, they found that teachers' styles tended to match the ideology of the school.

In a second study, Sternberg and Grigorenko (1995) examined style demographics for 124 students from four school sites: a large urban public high school, a college preparatory school, an elementary-secondary private school, and a small secondary parochial school. They found:

> ... that students were more positively evaluated and received better grades from teachers who matched their styles than from those who did not. Moreover, teachers tended to overestimate the extent to which their students matched them in styles. In other words, teachers think their students are more like them than they really are. (Sternberg, 1997a, p. 709)

The findings from this study seem to indicate that a mismatch between student-teacher cognitive styles may result in depressed student achievement.

Sternberg and Grigorenko (1997) also investigated the relationship between different types of abilities and different thinking styles. The concept of different styles of abilities is the cornerstone of Sternberg's triarchic theory of intelligence (Sternberg, 1988), in which he suggests that intellectual giftedness comes in three forms: analytic, creative, and practical. According to him, "the analytically gifted are strong in analyzing, evaluating, and critiquing; the creatively gifted are good at discovering, creating, and inventing; and

the practically gifted are strong in implementing, utilizing, and applying" (Grigorenko & Sternberg, 1997, p. 299). Sternberg and Grigorenko's study addressed two main questions. The first research question concerned the relationship between thinking styles (as measured by the TSI and the STSTS) and styles of ability. Second, the authors examined how well thinking styles predicted academic achievement (as measured by homework, final projects, and exams) after controlling for students' level of ability. Ability was measured by the Sternberg Triarchic Abilities Test (STAT; Sternberg, 1993). Third, different methods of instruction (analytical, creative, practical, and traditional) were utilized in the study to examine the effect that matching teaching and thinking styles had on student performance. Lastly, the authors investigated the relationship between students' thinking style and their performance on different types of evaluation methods (e.g., homework, written examinations, and a project).

Participants in the study were 199 high school students attending a summer program for gifted students. As part of the summer program, all students took part in a month-long psychology course. The afternoon session of the course was divided into four sections, and each section was taught using a different instructional type (memory, analytical thinking, creative thinking, or practical thinking). Some students were assigned to a classroom based on a match between their assessed thinking style and the instructional style, while others were purposely mismatched. All students received similar classroom assessments: two major homework assignments, a final project, and two exams. Each assessment was designed to test analytic, creative, and practical skills. Results showed that students' thinking styles did not differ across sex, grade, or ability patterns. Multiple

regression analyses were then conducted to assess whether thinking styles were predictive of academic performance after controlling for ability. In these analyses, academic performance (the dependent variable) had six levels: three assessment components (analytic, creative, and practical) and three assessment settings (homework, exams, and project). The two independent variables were abilities (STAT-analytical, STAT-creative, and STAT-practical), and thinking styles (the 13 subscale scores). Four of the 13 styles had significant ($p < .05$) positive correlations with the performance measures: Judicial, Legislative, Liberal, and Hierarchical (Appendix A). These correlations ranged from .14 to .23. Interestingly, the Executive style was the only one to have significant negative correlations with the performance measures (ranging from -.15 to -.18). That is, students who scored high on the Executive style tended to do poorly across all performance measures. The Judicial style was the only style to correlate significantly across all six performance variables (all positive correlations).

There was only one significant correlation between thinking style and ability: the STAT creative component and the Liberal-Conservative type correlated .22, with Liberal students being more creative. This dearth of significant correlations lends support to Sternberg's assertion that thinking style and ability style are unrelated.

A multiple regression analysis was then conducted using the same six dependent variables as before, but limiting the independent variables to the 3 STAT components and the 5 thinking styles that correlated significantly with measures of academic performance (Judicial, Legislative, Executive, Liberal, and Hierarchical). The independent variables accounted for 16% of the variance in the analytic tasks ($F = 5.3$, $p < .0001$). Performance

on analytic tasks was best predicted by the analytical component of the STAT, followed by the Judicial, Legislative, and Executive styles. The predictors explained 15% of the variance on the creative tasks ($F = 5.0$, $p < .0001$). Performance on creative tasks was best predicted by the creative component of the STAT, followed by the Judicial and Executive styles. Finally, the dependent variables accounted for 13% of the variance on the practical tasks ($F = 4.1$, $p < .001$, $B = .18$). Performance on the practical tasks was best predicted by the practical component of the STAT, followed by the Judicial style. In general, students' ability was the best predictor of performance on different types of tasks, but thinking style (particularly the Judicial, Legislative, and Executive styles) also contributed.

In terms of the various assessment settings (homework, project, and exams), the analysis showed that the analytic component of the STAT was the best predictor of performance across all assessment settings ($p < .0001$ across all three models, B ranging from .24 to .27). In general, students who performed higher on the analytic component of the STAT tended to perform better on the homework assignments, project, and exams. The Judicial style was also predictive of performance on all three assessments: exams, homework, and final project. Performance on the final projects was predicted by the Legislative, Liberal, and Executive styles.

No interaction effect was found between the type of instruction provided and the thinking styles of students. In other words, students who received instruction matched to their thinking style did not perform better than mismatched students. This finding is inconsistent with other studies of the match between thinking and instructional styles (e.g., Dunn, 1984; Dunn, Griggs, Olson, Beasley, & Gorman, 1995). Students were then

categorized according to their thinking style to determine if types of assessments differentially benefitted students of different styles. Using analysis of variance, Sternberg and Grigorenko found that highly Judicial, Liberal, and Oligarchic thinkers did better than all other types on exams, Judicial thinkers did better than others on homework, and Executive and Anarchic thinkers were superior on projects. In addition, certain combinations of styles seemed to be advantageous with certain assessment types (e.g., students who were Judicial and Global did better than others on exams).

In sum, results of their study revealed no differences in thinking style across sex, grade, or ability pattern for a group of gifted high school students. Certain thinking styles contributed significantly to the prediction of academic performance, regardless of the type of instruction that was given. Furthermore, students with certain thinking styles tended to do better in some forms of evaluation than in others.

Limitations of Studies on Cognitive Style and School Success

Despite these promising results, several limitations mediate the findings of their study. First, the cognitive ability and style of the participants was assessed through an unpublished test with little evidence supporting its psychometric properties. The validity of the variable "cognitive styles" is, therefore, in question. Second, the type of instruction that each of the four groups received was not adequately described. Was the analytic instruction really distinct from the instruction focused on creative thinking? How did the researchers control for variability in the four instructors? Third, and finally, the assessments (homework, projects, and exams) were not thoroughly defined, rendering it difficult to assess the type of thinking that each assignment evokes and the comparability

of the assignments across the four instructional groups. In addition, the assessment instruments used were teacher-made, and thus lacked evidence of adequate reliability and validity. The results would have been more convincing if published, well-supported tests had been used in lieu of classroom assessments. Notwithstanding these criticisms, the results of this study do provide some evidence that thinking style may, in fact, be a valid construct that contributes to students' success in academic tasks. Clearly, more research is needed to elucidate this construct and its implications for students across assessments.

Statement of the Problem

Performance assessments have been widely endorsed and adopted by educational agencies seeking an alternative or supplement to traditional testing instruments (e.g., standardized, multiple-choice exams). The most common type of performance assessment is the writing assessment, which is now being used in two-thirds of all state-wide testing programs (Council of Chief State School Officers, 1998). Advocates of performance assessments laud their authenticity and emphasis on measuring higher-order thinking skills through a format that is more consistent with students' natural learning (Frederikson, 1984; Wolf et al., 1991). In contrast, they decry the negative repercussions of standardized multiple-choice tests, including the indirect nature of these tests, their negative impact on the curriculum, and their emphasis on lower-order thinking (e.g., memorization of isolated facts and skills) (Frederikson & Collins, 1989). The growth of the performance assessment movement has been accompanied by a plethora of research examining the logistical (e.g., development, implementation) and psychometric (e.g., validity and reliability) aspects of performance assessments (Koretz et al., 1994; LeMahieu

et al., 1995). What has been largely ignored, however, is the question of how students' personal qualities or characteristics may differentially impact their success on performance assessments. For example, perhaps a student who is proficient and comfortable with the clear structure provided by a multiple-choice English exam testing knowledge of U.S. presidents would have difficulty when asked to present mastery of the same content in essay format. This student may prefer a high degree of structure and low ambiguity across all tasks, and is therefore adversely affected by the "freedom" provided by more ambiguous tasks such as performance assessments.

In his theory of cognitive style, Sternberg (1990) addressed this question in his contention that all students have preferences, or styles, that impact academic achievement. He argued that style has as strong an impact on achievement as does general cognitive ability (Sternberg, 1990). Sternberg (1997b) suggested that when the academic environment is mismatched with a student's cognitive style, the student may be adversely affected, and to the unknowing teacher or parent, may appear stupid or intransigent. Aspects of the academic environment that Sternberg has examined include the teacher's cognitive style, the form of instruction, and the types of class work and exams that are used. Sternberg and Grigorenko (1995) found that students whose style matched that of their teacher received better grades and were rated more positively than those whose style was mismatched. Grigorenko and Sternberg (1997) found no interaction between thinking style and the type of classroom instruction provided. Also, in a study examining the relationship between style and different academic performance components and settings, it was found that students with the legislative style consistently scored better on homework,

exams, and projects, regardless of the type of skill being measured (e.g., analytic, creative, or practical) (Grigorenko & Sternberg, 1997). Limitations of the aforementioned studies include a limited sample (e.g., restricted to high school students attending a summer program for gifted youth), the use of a measure of cognitive ability without substantiated validity and reliability (the STAT), and dependent variables that were not clearly defined and measured (e.g., "project").

The purpose of the current study was to further investigate whether students' cognitive styles impact their performance on two distinct types of tests: a standardized, multiple-choice test of language (Iowa Test of Basic Skills; ITBS), and a state-developed performance assessment of writing skill (FLORIDA WRITES!). The ITBS contains four language tests that "assess fundamental skills in using the conventions of standard written English" (Hoover et al., 1993, p. 88). Although not a direct measure of writing skill, the language component of the ITBS does purportedly measure students' proficiency in applying the conventional guidelines of spelling, capitalization, punctuation, and usage within an edit-and-revise format. The FLORIDA WRITES! test, on the other hand, is a performance assessment of students' writing skills. It requires students to respond, in writing, to assigned topics within a specified time period.

According to Sternberg's (1997b) theory, students with cognitive styles that are conducive to certain types of tests should have higher performance on tests that match their style than on those that are incongruent with their cognitive style. However, in order to accurately test this hypothesis it is necessary to examine the effects of a second variable that undoubtedly affects performance on achievement tests: general cognitive ability. In

general, scores on standardized tests of intelligence correlate .50 -.75 with achievement tests (Kamphaus, 1993). Therefore, students with higher levels of cognitive ability will, on average, perform better on achievement tests than students with lower levels.

Furthermore, Sternberg hypothesized that there is some relationship between cognitive ability as measured by conventional tests and cognitive style, particularly the Eexecutive style, although this hypothesis has not been examined with a test of cognitive ability other than the STAT (Sternberg, 1997b). He suggested that most current tests of general cognitive ability reward the Eexecutive style because the tests' items require the solution of prestructured problems, rather than the creation of one's own problems or the adjudication of existing problems (Sternberg, 1997b). Sternberg also contended that cognitive ability and cognitive style are two distinct, substantive constructs that both influence academic achievement. Therefore, in order to test the hypothesis that cognitive style contributes to prediction of students' performance on different types of achievement tests, it is necessary to collect data on students' general cognitive ability, which can then be controlled for statistically during the data analysis. Accounting for the variance in achievement that is due to variance in cognitive ability will thus yield a more accurate and valid estimate of the effect that cognitive style has on achievement test scores.

Sternberg (1994b) hypothesized that "conventional tests strongly favor test-takers with an executive, local style, and favor somewhat those with internal and conservative styles as well" (Sternberg, 1994b, p. 179). He described "conventional tests" as multiple-choice/short-answer tests (Sternberg, 1994b). Sternberg also asserted that depending on how they are scored, essay tests are biased in favor of certain styles. Tests such as the

FLORIDA WRITES! that emphasize the examinee's ability "to criticize and otherwise analyze the material learned," (p. 179) or "to go beyond the information learned...[to demonstrate]...some kind of higher-order synthesis" (p. 179) will benefit students with a judicial style. In addition, in a test like FLORIDA WRITES! where organization counts, the hierarchical student will also benefit.

The main hypothesis tested in this study, then, is whether the addition of information regarding cognitive style strengthens the prediction of students' scores on the ITBS and FLORIDA WRITES! over the prediction already made by cognitive ability. If, after controlling for cognitive ability, the addition of cognitive style contributes to the prediction of scores on the two achievement measures, then the strength of the relationships between each style and the dependent measures can be closer examined to see if, as Sternberg suggests (1994b), certain tests are biased in favor of particular styles.

To date, no research has examined the effect of cognitive style on standardized, multiple-choice items or on performance-based assessment. This study is of significant importance, then, for two reasons: one, it would initiate an important line of research investigating the role that personal variables may play in students' achievement on the relatively new but widely-acclaimed performance assessments. Second, it would serve as a validation study for Sternberg's theory of mental self-government and his instrument measuring students' cognitive styles, neither of which has yet been investigated by an outside researcher.

CHAPTER 2
METHODS

Participants

Participants in this study were 154 8th and 9th grade students enrolled in north central Florida public middle and high schools, ranging in age from 13 to 16 years (M = 13.8, SD = .49). Of the 154 participants, 74% (n = 114) were Caucasian, 21% (n = 32) African-American, 4% (n = 6) Hispanic, and 1% (n = 2) Asian. Forty-five percent (n = 69) were boys, and 55% (n = 85) girls. For inclusion in the study, participants must have taken both the Iowa Test of Basic Skills and the FLORIDA WRITES! tests in Alachua County within the 12 months prior to the study. Only those students that returned signed parental consent forms were allowed to participate. Because of state laws that exclude students receiving exceptional education from participating in statewide testing programs, parental consent forms were distributed only to students enrolled in a regular education curriculum.

Instruments

Thinking Style Inventory (TSI)

The TSI (see Appendix B) is a self-report measure on which students rate themselves on a 1-7 Likert scale for a number of behaviors that indicate preferred strategies and ways people use to solve problems, to carry out tasks and projects, and to make decisions. This inventory is based on Sternberg's theory of mental self-government, which suggests that people govern their everyday activities using various styles that have

been developed throughout the lifetime. The theory proposes 13 components that construct an individual's style. The Thinking Styles Inventory consists of 65 items, five for each of the 13 scales delineated in Sternberg's theory.

The TSI yields 13 scores, one for each of the scales. The scales have been shown to have adequate internal-consistency reliability, ranging from .56 to .88, with a median of .82 (Sternberg, 1997b). Results of several unpublished factor analyses of the TSI have been generally supportive of the structure of the instrument (Sternberg, 1997b). One of these analyses resulted in a five-factor model (Sternberg, 1997b). Factor 1, Adherence to Structure, contrasted the liberal and legislative scales with the conservative and executive scales. This suggests that liberal people tend to be legislative, and conservative people tend to be executive. Factor 2, Engagement, demonstrated a contrast between the oligarchic and the judicial scales, suggesting that people who are priority setters tend to be judicial thinkers. According to Sternberg, this factor was unexpected and not consistent with the model (Sternberg, 1997b). Scope, Factor 3, contrasted the external and internal scales. This factor was predicted and consistent with the model. Factor 4, Level, contrasted the local and global scales, and was also predicted and consistent with Sternberg's model. Factor 5, Distribution of Time, was not predicted by the model but was consistent with the model. Only the hierarchic scale loaded on this factor.

The relationship between the TSI and two other measures of style was investigated by Sternberg (1991). Results indicate several significant correlations between subscale scores on the TSI, Myers-Briggs Type Indicator, and Gregorc Style Delineator (see Tables 1 and 2).

Table 1

Correlation Coefficients of Thinking Style Inventory Subscale Scores with the Meyers-Briggs Type Indicator

Subscale	Introversion/ Extraversion	Intuitive/ Sensation	Feeling/ Thinking	Perception/ Judging
Legislative	-.12	.24**	-.10	.10
Executive	-.02	-.51***	-.24**	-.50***
Judicial	-.18	.02	-.16	-.23*
Global	.03	.25*	.16	.10
Local	-.04	-.28***	-.16	-.26**
Progressive[a]	-.12	.39***	.00	.22**
Conservative	-.01	-.48***	-.15	-.36***
Hierarchical	-.08	-.27***	-.25**	-.50***
Monarchic	-.01	-.22**	-.04	-.09
Oligarchic	-.12	-.21*	-.05	-.18**
Anarchic	-.18*	.16	.06	.19*
Internalizing	.08	.14	-.09	.07
Externalizing	-.32***	-.11	.03	-.05

Note: From MSG Thinking Styles Inventory Manual, (p. 39), by Sternberg and Wagner.
[a]The Progressive subscale was later renamed the liberal subscale.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 2

Correlation Coefficients of Thinking Style Inventory Subscale Scores with the Gregorc Style Delineator

Subscale	Concrete/ Sequential	Abstract/ Sequential	Concrete/ Random	Abstract/ Random
Legislative	-.13	-.10	-.04	.24**
Executive	.47***	.31	-.28***	-.47***
Judicial	.06	.13	-.06	-.10*
Global	-.15	-.15	.10	.18*
Local	.27**	.01	-.17*	-.14
Progressive[a]	-.21*	-.18	-.02	-.36***
Conservative	.43	.25	-.22**	-.44***
Hierarchical	.42***	.16	-.35***	-.25**
Monarchic	.20*	.03	-.12	-.13
Oligarchic	.14	.09	-.04	-.17*
Anarchic	-.09	-.17*	.06	.17
Internalizing	.00	-.12	-.07	.14
Externalizing	-.03	.04	.04	-.03

Note: From MSG Thinking Styles Inventory Manual, (p. 39), by Sternberg and Wagner.
Note: [a]The Progressive subscale was later renamed the Liberal subscale.
*$p < .05$. **$p < .01$. ***$p < .001$.

Three correlation coefficients higher than .50 were found between subscales of the TSI and the Myers-Briggs Type Inventory when both instruments were administered to laypeople. Results indicate that persons who scored high on the Executive subscale of the TSI tended to be typed as Sensation and Judging by the MBTI ($r = .51$ and $r = .50$, respectively). Those scoring high on the Hierarchical subscale also tended to typed as Judging by the MBTI ($r = .50$, $p < .001$).

To date, only one published study has reported the reliability and validity of the TSI. Zhang and Sachs (1997) administered the short version of the inventory to 88 post-secondary students enrolled in the University of Hong Kong. The mean age of the students was 27 years. All participants were pursuing an undergraduate or graduate degree in education. Estimates of internal consistency for the 13 scales were obtained, followed by an exploratory factor analysis of the intercorrelation matrix of the 13 scales. Analyses of variance were used to assess group differences in thinking styles based on the variables of sex, age, college class, college major, length of teaching experience, subject area taught, number of hobbies, residential location, travel experience, and parents' mean education. The Cronbach alpha estimates of internal consistency for the 13 scales ranged from .53 to .87, with a median of .70. The exploratory factor analysis resulted in the retention of three uncorrelated factors which together accounted for 65% of the variance in scores. The communalities of the 13 subscales ranged from .38 (Global) to .86 (Internal). The Executive, Conservative, and Oligarchic subscales had the highest factor loadings on Factor 1 (.87, .90, and .73, respectively). The Factor 2 loadings were highest for the Legislative (.77) and Internal (.80) subscales, and the External subscale had the highest

factor loading on Factor 3 (.87). This differs substantially from the five factor model found by Sternberg (1997b). However, the pattern of factor loadings are, in general, supportive of Sternberg's theory. The exceptions are the concepts of level (Global and Local) and forms (Hierarchical, Monarchic, Oligarchic, and Anarchic). This factor model did not differentiate between the Global and Local types (i.e., both subscales loaded highly on at least two factors). In addition, three of the four subscales measuring Form loaded highly on Factors 1 and 3.

Zhang and Sachs (1997) found mean group differences in thinking styles on age, sex, class, teaching experience, college major, and school subject taught. None of the findings challenged Sternberg's theory of the thinking styles, as he has made few assertions concerning the relationship between demographic variables and thinking styles.

Overall, Zhang and Sachs (1997) found mixed results regarding the TSI. They found moderate internal consistency estimates of the 13 TSI subscales. Results of factor analysis did not replicate Sternberg's (1997b) five factor model, although many of the factor loadings were consistent with his theory. Overall, the results suggest that further research is needed on the factor structure of the TSI.

Raven's Standard Progressive Matrices (SPM)

The measure of general cognitive ability used in this study is the SPM (Raven, Court, & Raven, 1983). The SPM was chosen as the measure of cognitive ability in this study for several reasons. First, its high *g*-loading suggests that it is a rather pure measure of general cognitive ability, and is therefore an adequate measure of students' level of intelligence (Jensen, 1998). The fact that it is a nonverbal measure of intelligence is of

importance when considering the types of a statistical analyses that will be used to examine the data. If controlling for general cognitive ability is integral to the hypothesized model, then what is being stated is that cognitive ability is related to performance on the achievement tests, and that in order to remove the variance in achievement scores associated intelligence, scores on the intelligence test must be controlled for statistically. If a group measure of intelligence was chosen that contained a high verbal component (e.g., the Otis-Lennon), then controlling for scores on the intelligence test may remove excessive variance from scores on verbally-loaded achievement tests. Therefore, a better option would be to use a measure of g that does not overlap significantly with verbally-loaded achievement tests. Using the SPM as the measure of intelligence in this study is a defensible choice: it has adequate reliability and validity, it is highly g-loaded, and accounts for the relationship between intelligence and achievement without compromising the model.

The SPM is a nonverbal paper-and-pencil test of general cognitive ability. Raven's matrices test has three separate forms: The Coloured Progressive Matrices (for ages 5 to 11 years and mentally impaired adults), the Standard Progressive Matrices (ages 8 to 65 years), and the Advanced Progressive Matrices (for high-ability adolescents and adults). The Standard Progressive Matrices (SPM) was used with this sample based on the participants' age range and estimated cognitive ability.

The SPM form contains 60 items, presented in five sets, with 12 items per set. The sets are arranged in order of progressive difficulty. Each item is printed on an 8" x 10" page, and consists of a matrix-like arrangement of figural symbols. Subjects select the

appropriate missing symbol from a group of six symbols displayed at the bottom of the page, and record the number of the correct response (1-6) on the answer sheet.

Scores on the SPM have been shown to have adequate reliability and validity as an estimate of general cognitive ability. The split-half reliability is estimated to be .86 (Raven, Court, & Raven, 1983). The concurrent validity of the SPM has been established with numerous intelligence and achievement tests. Validity coefficients with intelligence tests are in the .50s to .80s, and tend to be higher with performance than with verbal tests (Anastasi, 1988; Sattler, 1992). Correlations with achievement tests are in the .30s to .60s (Sattler, 1992). Results from factor analyses suggest that although the SPM forms are heavily loaded with the psychometric g-factor, spatial aptitude, inductive reasoning, perceptual accuracy, and other factors also influence performance (Anastasi, 1988; Burke, 1958; Corman & Budoff, 1974; Sattler, 1988). Jensen (1980), however, states that when psychometric artifacts are accounted for statistically, the SPM measures g and very little else.

Iowa Test of Basic Skills (ITBS)

The ITBS is a battery consisting of standardized tests in several subject areas, including reading, math, and language. The ITBS is administered every spring to all 4th and 8th grade students in regular education classes. All items on the ITBS are of the multiple-choice format. Scores on the ITBS are norm-referenced, and are presented as standard scores, percentile ranks, stanines, and normal curve equivalents. Scores on the ITBS are used as general indicators of how well individual students, schools, counties, and states are performing as compared to others taking the same test. Scores are not used to

assign grades, determine eligibility for special services, or promote students to the next grade level.

Participants in this study were administered Level 14 of the ITBS Complete Battery, Form K in March of 1998. Scores of interest are those for Total Language, which is comprised of the four subtests of Spelling, Capitalization, Punctuation, and Usage and Expression. Appendix C lists the number of items and testing time allotted to each of these four tests. According to the ITBS Technical Manual (1994, Riverside Publishing Company), the language tests in the ITBS assess basic skills common to standard written English.

> The items used to measure spelling, capitalization, punctuation, and usage and expression sample developing skills that are of critical importance in the revising and editing stages of the writing process. Although it is recognized that a complex cognitive process like writing is not completely amenable to objective measurement, the goal of that process, the precise expression of an author's thoughts and feelings, entails a command of certain aspects of written language that can be effectively measured with objective test items....In levels 9-14 of the battery, there are four separate Language tests in the areas previously mentioned. Although the functional use of written language during revising and editing demands the application of skill in all of these areas simultaneously, sound measurement practice is still better served by the use of a separate test in each area. There are several reasons for this. Separate tests allow for more complete sampling of the content domains in spelling, capitalization, punctuation, and usage and expression than would a unitary test. (p. 21)

The ITBS was mostly recently standardized in 1992 on sample of approximately 136,000 students, stratified according to school type, socioeconomic status, and geographic region. The internal consistency reliability (KR 20) estimate for the eighth-grade students taking Level 14 of the ITBS was .95 for Total Language, ranging from .82 - .88 for each of the four subtests. The Technical Manual provides information regarding p -values for the

items on the ITBS, which ranged from the low .20s to the high .90s. Information regarding floor and ceiling effects is also provided.

FLORIDA WRITES! Assessment

The FLORIDA WRITES! test is a direct writing assessment requiring students to respond to prompts that are intended to elicit writing for specific purposes. The FLORIDA WRITES! assessment program was authorized in 1990 when the Florida Legislature revised Section 229.57 of the Florida Statutes (Florida Statewide Assessment Program, 1995). This writing assessment program was based upon teaching theory and standards developed by the Department of Education (DOE) in conjunction with school and district educators across the state of Florida. Advisory committees consisting of teachers, curriculum supervisors, school administrators, district testing supervisors, and citizens of different geographical areas, gender, racial, and ethnic groups assisted the DOE in the development of the program. Currently, scores on the FLORIDA WRITES! test are used to identify trends in writing over a period of several years and to identify strengths and weaknesses in various programs of writing instruction (Hoover et al., 1993). Scores are not used to assign grades, promote students, or determine eligibility for special services.

Currently, FLORIDA WRITES! is administered annually to students in grades 4, 8, and 10. Students that may be exempt from testing include:

> (a) students whose primary language is not English and who have been receiving services in an approved Limited English Proficiency (LEP) program for less than two years, (b) students who have a temporary physical disability or temporary emotional problem, or (c) full- or part-time exceptional education students with a current individual educational plan (IEP) who have been classified according to

> State Board of Education Rule 6A-6.331, FAC. (Florida Statewide Assessment Program, 1995, p. 9).

Fourth-graders are given either a narrative prompt or an expository prompt. Students in grades 8 and 10 are given an expository or a persuasive prompt. Narration is defined as:

> writing that tells about a personal or fictional experience or tells a story based on a real or imagined event. Narrative writing is characterized, as appropriate, by insight, creativity, drama, suspense, humor, or fantasy. The unmistakable purpose of this type of writing is to create a central theme or impression in the reader's mind. (Florida Statewide Assessment Program, p. 3)

Exposition is:

> writing that gives information, explains why or how, clarifies a process, or defines a concept. Though objective and not dependent on emotion, expository writing may be lively, engaging, and reflective of the writer's underlying commitment to the topic. The unmistakable purpose of this type of writing is to inform, clarify, explain, define, or instruct. (Florida Statewide Assessment Program, p. 3)

The third type of prompt, persuasion, is described as:

> writing that attempts to convince the reader that a point of view is valid or that the reader should take a specific action. If it is important to present other sides of an issue, the writer does so, but in a way that makes her or his position clear. The unmistakable purpose of this writing is to convince the reader of something. (Florida Statewide Assessment Program, p. 5)

Each item (called a "probe") consists of two components: the Writing Situation and the Directions for Writing. The Writing Situation provides the topic or theme of the writing intended to "provide a common understanding of the meaning of the theme by expanding, restating, or clarifying the central idea for the student" (Florida Statewide Assessment Program, p. 5). An example is, "The principal at your school has been asked to discuss with a parent group the effect of watching television on students' grades" (Florida

Statewide Assessment Program, p. 5). The Directions for Writing provide a strategy intended to help those students who may have difficulty in getting started. An example for a persuasion probe is "Before you begin writing, think about the effect watching television has on your grades and your friends' grades. Now write to convince your principal to accept your point of view on the effect watching television has on grades" (Florida Statewide Assessment Program, p. 5).

Each student receives a writing folder which consists of one writing prompt with two lined pages for the written response. Students are given 45 minutes to read the prompt, plan their response, and to write their response in the folder. Responses are scored by trained readers using the holistic method to evaluate the response for its overall quality. The score is based on consideration of four elements: focus, organization, support, and conventions. Focus refers to how the writer maintains a main idea, theme, or unifying point. Organization refers to the structure of the response (beginning, middle, and end) and whether the writer's points are logically related. In addition, organization also refers to the use of transitional devices and connection between sentences. Support is concerned with the quality of details that the writer uses to explain, clarify, or define points. This includes word choice, specificity, depth, and thoroughness. Finally, convention refers to punctuation, capitalization, spelling, and variation in sentence structure (see Appendix D for complete holistic rubric). Readers award each written response a single score, ranging from 1 (lowest) to 6 (highest). See Appendix E for an example of the eighth-grade scoring rubric.

Validity of the FLORIDA WRITES! prompts was ensured in several ways. First, research on the writing process and issues in designing a direct writing assessment were examined. Advisory committees developed instruction on the type of writing (e.g., expository, narrative, and persuasive), cue words, writing situations, and directions for writing that were to be included in each prompt. Prompts were submitted by committee members and DOE staff, taken from other states' testing programs, and purchased from an educational consortium. Following pilot test administration, each prompt was reviewed by the advisory committee according to the guidelines and then field-tested in Florida school districts representing typical achievement levels of Florida's fourth, eighth, and tenth-grade students. Prompts were analyzed statistically and by expert judgements. Reviewers examined the word choice, syntax, concept familiarity, sentence complexity, and overall reading level of each prompt. In addition, prompts were reviewed to help prevent bias for or against any student because of gender, race, ethnicity, or geographical region. Prompts that met the criteria established by the advisory committee and DOE staff were accepted into the FLORIDA WRITES! item pool.

Responses are scored in North Carolina and Florida due to the large number of students. There are two scoring directors per grade level (hired by the scoring contractor), which in turn trained the team leaders, who subsequently trained the readers. Readers were required to have a bachelor's degree, and they had to submit a piece of their own writing that was evaluated by the scoring contractor. During training, committee-scored responses were given to potential readers who had to score a minimum of 70% of the

papers with exact agreement to qualify. Training of readers continues before and during scoring.

Each student response is read by two readers. If the scores are in perfect agreement or have contiguous agreement, then the average of these two scores is assigned. Noncontiguous scores are read by a team leader who subsequently assigns a score. The scoring process is continuously monitored by checking ratings, the number of split ratings, and the distribution of ratings for each rater. For the 1995 assessment, the percentage of perfect agreement ranged from 57.4% to 63.7% across grade and type of writing. The percentage of noncontiguous scores ranged from 1.9% to 4.2%, and the percentage of scores within one score point ranged from 95.8% to 98.1% across grade and type of writing. Table 3 represents the interrater reliability by prompt for the 1995 eight-grade assessment.

Data from the 1995 FLORIDA WRITES! test administration show that 3.0 was the most frequently awarded score for both the expository and persuasive prompts. For the expository prompt, 85.35% of the eighth-grade students received a score between 2.0 and 4.0 ($\underline{N}$ = 62,946). For the persuasive prompt, 84.5% received a score between 2.0 and 4.0 ($\underline{N}$ = 62, 508). See Appendix F for more specific data on the distribution of scores across prompts for the eighth-graders.

Table 3

Interrater Reliability by Prompt for the Eighth-Grade Scores in 1995

	Prompt			
	Expository		Persuasive	
	Number	Percentage	Number	Percentage
Number Read	62, 946		62, 508	
Perfect Agreement	36,310	57.7	36,743	58.8
Contigous Agreement	23,989	38.1	23,830	38.1
Noncontigous Scores	2,647	4.2	1,935	3.1

Source: Florida Statewide Assessment Program (1995). Florida Writes! 1995 Technical Report, p. 20.

Procedure

Approval of the study was obtained by the Institutional Review Board (IRB) of the University of Florida and the Alachua County School Board. Application for Research in Alachua County forms were sent to all middle and high school schools in the county during two phases. First, all middle schools were contacted in Spring of 1998 regarding the participation of eighth-grade students. High schools were contacted in the Fall of 1998 regarding ninth-grade students. Consent was granted by four middle and four high schools. Parental consent forms (Appendix G) were then distributed to all regular education classrooms within the participating schools. Students in Exceptional Student Education classrooms were excluded from this study because at that time, many of these students were excluded from the ITBS and FLORIDA WRITES! assessment programs.

Following a period of two weeks, the parental consent forms were collected and testing times and locations were arranged with school personnel. All students returning signed parental consent forms were allowed to participate in the study regardless of sex, race, ethnicity, religion, and primary language.

All testing took place during school hours. During the scheduled testing time at each school, participants were gathered in a vacant classroom and seated at individual desks. The examiner read the student assent form (Appendix H) aloud and gave the participants the opportunity to leave if they desired. The following materials were then distributed to each participant: the SPM test booklet and response sheet, the TSI and response sheet, and one pencil. Researchers adhered to the standardized instructions for administering the SPM and the TSI as outlined in their respective manuals. Standardized instructions for both instruments were read aloud. Participants were given unlimited time to complete the instruments. As each participant completed the instruments, they were allowed to return to their classroom. On average, the entire testing procedure lasted 1 ½ hours.

Following data collection, all protocols were scored by the principal investigator. Student's scores on the SPM and TSI were matched to their scores on the ITBS and FLORIDA WRITES!, obtained from the Alachua County data base. After matching, all scores were added to a new data base which replaced student names with code numbers. This new data base was used for statistical analyses.

Analysis

The main hypothesis of whether the addition of cognitive style contributes to the prediction of student scores' on the ITBS and FLORIDA WRITES! will be tested using multiple regression. In separate analyses, scores on the ITBS and FLORIDA WRITES! will serve as dependent variables. The salience of cognitive ability as a predictor of ITBS will be examined first, using the SPM as the measure of cognitive ability. The subscales within each of the five domains of the TSI will then be simultaneously entered, resulting in five separate regression equations. For example, the equation for testing the domain of Function will contain the following independent variables: SPM, Executive, Local, and Judicial scores. For each equation, the increase in the predictive ability of the full model (i.e., SPM plus TSI subscales) over the reduced model (i.e., SPM alone) will be examined for significance. These analyses will reveal if, after controlling for cognitive ability, cognitive style predicts ITBS scores. A second set of analyses will follow the same procedure, using FLORIDA WRITES! as the dependent measure. The reduced model will contain both SPM and ITBS scores as independent variables. The full model will contain SPM, ITBS, and the subscale scores for each domain, resulting in five separate regression equations. Again, the increase in the predictive ability of the full model over the reduced model will be the focus of these analyses. If any domain of the TSI is found to predict achievement scores, post-hoc analyses will be conducted to test the strength of the regression coefficients between the subscale scores and the achievement scores.

CHAPTER 3
RESULTS

Review of Analyses

Tables 4, 5, and 6 present descriptive statistics for the variables across all participants. Raw scores (ranging from 1 to 6) for FLORIDA WRITES! were used. For the ITBS, national curve equivalent (NCE) scores were computed from percentiles. NCE scores are standard scores with a mean of 50 and a standard deviation of 21.06, and range from 1 to 99. In comparison to percentile ranks, NCE scores are more spread out at the extremes and less spread out near the middle. SPM data were presented as raw scores (ranging from 0 to 60). Table 4 presents descriptive statistics for the variables of FLORIDA WRITES!, ITBS, and SPM.

Table 4

Descriptive Statistics for Tests of Achievement and Cognitive Ability

Variable	M	SD	N
FLORIDA WRITES!	3.5	.75	144
Iowa Test of Basic Skills[a]	59.1	19.0	146
Raven's SPM[b]	47.0	6.2	152

[a]National Curve Equivalent (NCE) scores were used.
[b]SPM = Standard Progressive Matrices.

The mean FLORIDA WRITES! ($M = 3.5$) score for this group of participants is above the state average (3.0) and falls at approximately the 76th percentile (see Appendix F for scoring distributions across the state). The mean ITBS score of 59.1 falls above the national mean, at approximately the 66th percentile. These results suggest that the achievement of this sample of students falls at levels slightly above the state and national means.

Table 5 presents the descriptive data for FLORIDA WRITES!, ITBS, and SPM by sex. Homogeneity of variance was tested for all three variables using Leven's Test for Equality of Variances. Equal variances between boys and girls were found for FLORIDA WRITES! and SPM scores ($p > .05$). For the ITBS, however, girls' scores were significantly more spread out than boys' ($F = 9.2$, $p < .05$). Assuming equal variances, a t-test for independent samples was conducted for the variables of FLORIDA WRITES! and SPM. Boys' and girls' scores did not differ significantly on SPM, $t(149) = -1.3$, $p > .05$. On FLORIDA WRITES!, however, girls scored significantly higher than boys, $t(142) = -4.0$, $p < .001$. Accomodating for the unequal variances, a t-test for independent samples was then conducted for the ITBS scores. In this sample, girls performed significantly better on the ITBS than did boys, $t(137) = -2.1$, $p < .05$.

Table 5

Achievement and Ability Test Scores for Boys and Girls

	Males			Females		
Variable	M	SD	N	M	SD	N
FLORIDA WRITES!	3.2	.63	66	3.7	.78	78
Iowa Test of Basic Skills[a]	55.7	14.9	68	62.0	21.7	78
Raven's SPM[b]	46.5	6.3	71	47.8	5.8	80

[a]National Curve Equivalent (NCE) scores were used.
[b]SPM = Standard Progressive Matrices.

Descriptive statistics for the 13 subscales of the TSI are presented in Table 6. Subscale scores for the TSI are presented as raw scores, with a possible range of 0 to 35. The mean scores represent the group mean Likert ratings for each subscale. The subscale with the highest mean endorsement was the Judicial subscale ($M = 25.8$). The Global subscale received the lowest overall endorsement ($M = 20.4$). An independent samples t-test was used to examine differences between boys' and girls' endorsements on the TSI subscales. Homogeneity of variance was demonstrated for all subscales. Two of the thirteen subscales showed higher means for girls than boys. The girls' mean endorsement ($M = 25.0$, $SD = 4.8$) for the Anarchic subscale was higher than the boys' ($M = 22.8$, $SD = 5.4$), $t(151) = -2.8$, $p < .01$. On the External subscale, the girls' mean score ($M = 24.8$, $SD = 6.4$) was also higher than the boys ($M = 22.2$, $SD = 6.2$), $t(151) = -2.6$, $p < .05$. The mean ratings for the remaining 11 subscales did not differ across sex.

Table 6

Descriptive Statistics for Thinking Styles Inventory Subscales

Variable	M	SD	N
Executive	21.1	5.9	154
Legislative	25.8	6.4	154
Judicial	21.5	5.4	154
Global	20.4	4.5	154
Local	20.7	4.8	154
Liberal	23.0	6.2	154
Conservative	20.5	5.1	153
Hierarchic	23.1	5.1	154
Monarchic	22.8	4.9	154
Oligarchic	21.0	5.5	154
Anarchic[a]	23.9	5.2	154
Internal	22.6	6.0	154
External[b]	23.6	6.4	154

[a]Girls' endorsement on this subscale signficantly higher than boys', $p < .01$.
[b]Girls' endorsement on this subscale signficantly higher than boys', $p < .05$.

Internal Consistency

Estimates of internal consistency for the 13 subscales of the TSI were obtained. Table 7 lists the Cronbach alpha (α) estimates for each of the 13 subscales.

Table 7

Cronbach's Alpha (α) Estimates for the Thinking Styles Inventory Subscales (N=153)

Scale	Items	α
Legislative	5,10,14,32,49	.62
Executive	8,11,12,31,39	.75
Judicial	20,23,42,51,57	.68
Global	7,18,38,48,61	.51
Local	1,6,24,44,62	.56
Liberal	45,53,58,64,65	.81
Conservative	13,22,26,28,36	.65
Hierarchic	4,19,33,25,56	.64
Monarchic	2,43,50,54,60	.48
Oligarchic	27,29,30,52,59	.68
Anarchic	16,21,35,40,47	.63
Internal	9,15,37,55,63	.70
External	3,17,34,41,46	.78

Cronbach alpha estimates ranged from .48 (Monarchic subscale) to .81 (Liberal), with a median of .65. These estimates were lower than those reported in other studies of the reliability of the TSI. Sternberg (1997b) reported a range of .57 to .88, with a median of .82. Zhang and Sachs (1997) reported a range of .53 to .87, with a median of .70.

Multiple Regression

Review of Procedure for Testing Main Hypothesis

Multiple regression was used to test whether cognitive style contributed to the prediction of students' scores on the ITBS and FLORIDA WRITES!. Scores on the SPM were included to control for the variance in achievement test scores attributed to cognitive ability. The full correlation matrix for all variables is displayed in Appendix I. The first step in the analysis began by estimating in one equation the ability of SPM scores to predict ITBS scores. In a second equation, the subscale scores of the TSI Function domain (i.e., Executive, Legislative and Judicial) were added to the model, resulting in four independent variables: SPM, Executive, Legislative, and Judicial. The increase in the proportion of the variance explained by the full model over the reduced model indicated whether the Function styles contributed significantly to the prediction of ITBS scores after controlling for cognitive ability. A comparison of the full and reduced models was made for each domain of the TSI (e.g., Function, Level, Leaning, Form, and Scope). The same procedure was then repeated, using FLORIDA WRITES! as the dependent variable. ITBS scores were included in this second set of analyses as an independent variable in order to examine for the variance in FLORIDA WRITES! related to ITBS.

Assumptions

Before conducting the analyses, it was important to assess whether the data met the assumptions for multiple regression. Assumptions for multiple regression generally fall into two categories: (a) those involving residual (error) scores, and (b) those involving measurement errors (Licht, 1995). A residual is the difference between a subject's actual score on the criterion and the predicted score based on the regression equation. Residuals should be homoscedastic (have equal variances) and be normally distributed. The normality assumption was tested using the ZRESID command of SPSS, which plots the observed cumulative probabilities of the residual scores against the expected cumulative probabilities. The plots suggested that the residuals for all of the multiple regression analyses were normally distributed. The scatterplot of predicted scores against residuals showed no pattern, confirming that the assumptions of linearity and homoscedasticity were met.

The primary concern of this study in regards to measurement error was multicollinearity. Multicollinearity occurs when the intercorrelations between the independent variables are too high, suggesting that the variables may be measuring the same construct. In this study, multicollinearity was examined by computing the squared multiple correlation (R^2) for the independent variables used in each of the multiple regression analyses. The R^2 s ranged from .00 to .33. Tolerance levels were then computed by subtracting each R^2 from 1.0. Tolerance levels less than .10 indicate problems of collinearity (J. Algina, personal communication, October, 1996). The highest

Tolerance level found in this study was .67. Results of these analyses, therefore, demonstrate an absence of multicollinearity.

Findings

Results of the multiple regression analyses did not support Sternberg's (1997b) hypothesis that thinking style contributes to the prediction of students' performance on achievement tests. After controlling for cognitive ability, scores on the TSI subscales were not related to performance on either the ITBS or FLORIDA WRITES!.

The first set of analyses included ITBS as the dependent variable. The first step was to build a regression model with SPM as the sole independent variable. Thus, the first estimated regression model was: $\underline{Y} = \underline{a} + \underline{b}_1 \underline{x}_1 + \underline{e}$, where $\underline{Y}$ represents the predicted value of the criterion, $\underline{a}$ represents the y-intercept, $\underline{b}_1$ is the regression estimate for the first independent variable $\underline{x}_1$, and $\underline{e}$ is the error term. Substituting the variable names for symbols, the estimated regression equation was: ITBS = $\underline{a}$ + $\underline{b}_1$SPM. This reduced model was later compared with a full model that included the TSI variables. The summary ANOVA table for the reduced model is presented in Table 8.

The $\underline{R}^2$ represents the proportion of variance in the criterion (e.g., ITBS) that is accounted for by the variance of the combination of predictor variables. The Adjusted $\underline{R}^2$ was used in all multiple regression analyses because the uncorrected $\underline{R}^2$ is a biased estimate of the population parameter. Table 8 shows that the Adusted $\underline{R}^2$ for SPM regressed onto ITBS was .26. Therefore, 26% of the variance in ITBS scores is accounted for, or shared with, variance in SPM scores. This is considered a statistically significant proportion ($\underline{p} < .001$).

Table 8

Summary ANOVA Table for Reduced Model: Cognitive Ability (SPM) Regressed on Iowa Test of Basic Skills

Source of Variance	df	SS	MS	F	R^{2a}
Model	1	13461.57	13461.57	49.67	.26*
Error	141	38214.87	271.03		

[a]Adjusted R^2 were used. $R^2 = .26$.
$p < .001$.

After building the reduced model, five separate full models were tested, one for each domain of the TSI (viz., Function, Level, Leaning, Form, Scope). Each estimated regression model followed the format $Y = a + b_1 x_1 \ldots b_g x_g + e$, where b_g is the regression estimate for the last subscale, x_g, within the domain being tested. For example, the full model for the domain of Function was: ITBS = $a + b_1$ SPM + b_2Executive + b_3 Legislative + b_4 Judicial + e. The summary ANOVA table for Function is in Table 9. The regression results for the domain of Function are displayed in Table 10.

Table 9

Summary ANOVA Table for Full Model: Cognitive Ability (SPM) and Function Subscales (Executive, Legislative, Judicial) Regressed on Iowa Test of Basic Skills

Source of Variance	df	SS	MS	F	R^{2a}
Model	4	14056.77	3514.19	12.89	.25*
Error	137	37359.83	272.70		

[a]Adjusted R^2 were used. $R^2 = .27$.
$p < .001$.

Table 10

Results: Cognitive Ability (SPM) and Function (Executive, Legislative, Judicial) Subscales Regressed on Iowa Test of Basic Skills

	Regression	Standard	Standardized
Variable	Coefficient	Error	Regression Coefficient
SPM[a]	1.55	.23	.50*
Executive	-.01	.27	-.00
Legislative	.19	.23	.07
Judicial	.17	.31	.05

[a]SPM = Standard Progressive Matrices.
$p < .001$.

The Adjusted R^2 for the full model indicates that 25% of the variance in ITBS scores is accounted for by the combined predictors of SPM and the Executive, Legislative, and Judicial subscales of the TSI. This was compared to the Adjusted R^2 from the reduced model, which was also 25%. Clearly, the addition of subscale scores from the Function domain of the TSI did not explain more variance of the ITBS than did the SPM scores. In other words, cognitive ability was the best predictor of ITBS scores; whether a student had an Executive, Legislative, or Judicial style was not a factor that influenced performance on the ITBS.

In this first comparison, a statistical test was not necessary to discern a statistical difference in the R^2's between the full and reduced models because the two R^2s were equal (.25). However, in subsequent comparisons, the following formula was used to compute an F ratio for the difference in R^2s between the reduced and full models:

$$F = \frac{n - k - 1}{k - g} \cdot \frac{R^2(FM) - R^2(RM)}{1 - R^2(FM)}$$

The n represents the number of subjects, k is the number of b's in the full model, and g is the number of b's in the reduced model. Thus, in the analysis of the Function domain (i.e., Executive, Legislative, and Judicial subscales), $k = 4$ and $g = 1$. The $R^2(FM)$ denotes the squared multiple correlation for the full model. The $R^2(RM)$ is for the reduced model. The critical value used was $F_{.05,\, k-g,\, n-k-1}$.

Tables 11 through 18 present data from the hierarchical analyses and the summary ANOVA tables for the remaining four domains of the TSI: Level, Leaning, Form, and Scope. Each full model was based on one of these domains. The results of the full models

were compared individually to the reduced model presented previously ($R^2 = .25$) to examine if inclusion of that domain in the model would contribute more to the prediction of ITBS than SPM alone. Each summary ANOVA table includes the R^2 for that full model. Included in each regression table are the beta coefficients, standard errors, and standardized beta coefficients. Below each regression table are the F ratios and critical values for F representing the difference in R^2s between the full and reduced models.

Table 11

Summary ANOVA Table for Full Model: Cognitive Ability (SPM) and Level (Global, Local) Subscales Regressed on Iowa Test of Basic Skills

Source of Variance	df	SS	MS	F	R^{2a}
Model	3	13758.47	4586.16	16.81	.25*
Error	138	37658.13	272.89		

[a]Adjusted R^2 were used. $R^2 = .27$.
$p < .001$.

Table 12

Results: Cognitive Ability (SPM) and Level (Global, Local) Subscales Regressed on Iowa Test of Basic Skills

Variable	Regression Coefficient	Standard Error	Standardized Regression Coefficient
SPM[a]	1.58	.23	.52*
Global	.14	.32	.03
Local	.12	.31	.03

Note. For increase in R^2, $F = -.27$, $F_{c.v.} \approx 3.0$.
[a]SPM = Standard Progressive Matrices.
$p < .001$.

Table 13

Summary ANOVA Table for Full Model: Cognitive Ability (SPM) and Leaning (Liberal, Conservative) Subscales Regressed on Iowa Test of Basic Skills

Source of Variance	df	SS	MS	F	R^{2a}
Model	3	13769.53	4589.84	16.86	.25*
Error	138	37290.21	272.19		

Note. For increase in R^2, $F = -.09$, $F_{c.v.} \approx 3.00$.
[a]Adjusted R^2 were used. $R^2 = .27$.
$p < .001$.

Table 14

Results: Cognitive Ability (SPM) and Leaning (Liberal, Conservative) Subscales Regressed on Iowa Test of Basic Skills

Variable	Regression Coefficient	Standard Error	Standardized Regression Coefficient
SPM[a]	1.62	.23	.53*
Liberal	-.26	.22	-.09
Conservative	.00	.27	-.01

[a]SPM = Standard Progressive Matrices.
$p < .001$.

Table 15

Summary ANOVA Table for Full Model: Cognitive Ability (SPM) and Form (Hierarchic, Monarchic, Oligarchic, Anarchic) Subscales Regressed on Iowa Test of Basic Skills

Source of Variance	df	SS	MS	F	R^{2a}
Model	5	14431.48	2886.30	10.61	.25*
Error	136	36985.12	271.95		

Note. For increase in R^2, $F = .05$, $F_{c.v.} \approx 2.37$.
[a]Adjusted R^2 were used. $R^2 = .28$.
$p < .001$.

Table 16

Results: Cognitive Ability (SPM) and Form (Hierarchic, Monarchic, Oligarchic, Anarchic) Subscales Regressed on Iowa Test of Basic Skills

Variable	Regression Coefficient	Standard Error	Standardized Regression Coefficient
SPM[a]	1.58	.23	.51*
Hierarchic	.47	.34	.13
Monarchic	-.31	.33	-.08
Oligarchic	-.01	.27	-.02
Anarchic	-.33	.31	-.09

[a]SPM = Standard Progressive Matrices.
$p < .001$.

Table 17

Summary ANOVA Table for Full Model: Cognitive Ability (SPM) and Scope (Internal, External) Subscales Regressed on Iowa Test of Basic Skills

Source of Variance	df	SS	MS	F	R^{2a}
Model	3	14412.14	4804.05	17.92	.27*
Error	138	37004.45	268.15		

Note. For increase in R^2, $F = .93$, $F_{c.v.} \approx 3.00$.
[a]Adjusted R^2 were used. $R^2 = .28$.
$p < .001$.

Table 18

Results: Cognitive Ability (SPM) and Scope (Internal, External) Subscales Regressed on Iowa Test of Basic Skills

Variable	Regression Coefficient	Standard Error	Standardized Regression Coefficient
SPM[a]	1.54	.23	.50*
Internal	.39	.24	.12
External	-.01	.23	-.02

[a]SPM = Standard Progressive Matrices.
*$p < .001$.

As the data in Tables 9 through 18 demonstrate, the addition of the TSI subscales to the SPM did not increase prediction of students' ITBS scores. Cognitive ability was the only variable that explained a significant proportion of the variance in ITBS scores. Cognitive style did not impact students' achievement scores.

The second set of multiple regression analyses used FLORIDA WRITES! scores as the dependent variable instead of ITBS. In the reduced model, the predictor variables were both SPM and ITBS ($g = 2$). Substituting variable names, the reduced model was FLORIDA WRITES! = $a + b_1$SPM + b_2 ITBS. Tables 19 and 20 present the summary ANOVA and regression tables for this model.

Table 19

Summary ANOVA Table for Reduced Model: Cognitive Ability (SPM) and Iowa Test of Basic Skills Regressed on FLORIDA WRITES!

Source of Variance	df	SS	MS	F	R^{2a}
Model	2	20.15	10.07	23.96	.25*
Error	137	57.59	.42		

[a]Adjusted R^2 were used. $R^2 = .26$.
$p < .001$.

Table 20

Results of Reduced Model: Cognitive Ability (SPM) and Iowa Test of Basic Skills Regressed on FLORIDA WRITES!

Variable	Regression Coefficient	Standard Error	Standardized Regression Coefficient
SPM[a]	.00	.01	-.03
ITBS[b]	.02	.00	.52*

[a]SPM = Standard Progressive Matrices. [b]ITBS = Iowa Test of Basic Skills.
$p < .001$.

Results of this analysis demonstrate that together, cognitive ability and performance on the ITBS accounted for 25% of the variance in FLORIDA WRITES!. It is also noted that after controlling for SPM, ITBS correlates significantly with FLORIDA WRITES!,

suggesting the overlap between ITBS and FLORIDA WRITES! is not simply cognitive ability.

Like before, full models for each domain were then compared to the reduced model to examine if cognitive style contributed to the prediction of FLORIDA WRITES! made by cognitive ability and ITBS. Tables 21 through 30 display the results of these analyses.

Table 21

Summary ANOVA Table for Full Model: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Function (Executive, Legislative, Judicial) Subscales Regressed on FLORIDA WRITES!

Source of Variance	df	SS	MS	F	R^{2a}
Model	5	21.26	4.25	10.01	.25*
Error	138	56.48	.43		

Note. For increase in R^2, $F = -.12$, $F_{c.v.} \approx 2.60$.
[a]Adjusted R^2 were used. $R^2 = .27$.
$p < .001$.

Table 22

Results: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Function (Executive, Legislative, Judicial) Subscales Regressed on FLORIDA WRITES!

Variable	Regression Coefficient	Standard Error	Standardized Regression Coefficient
SPM[a]	.00	.01	-.02
ITBS[b]	.02	.00	.52*
Executive	.01	.01	.10
Legislative	.00	.01	-.02
Judicial	.00	.01	.03

[a]SPM = Standard Progressive Matrices. [b]ITBS = Iowa Test of Basic Skills.
$p < .001$.

Table 23

Summary ANOVA Table for Full Model: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Level (Global, Local) Subscales Regressed on FLORIDA WRITES!

Source of Variance	df	SS	MS	F	R^{2a}
Model	4	21.26	5.32	12.61	.25*
Error	134	56.47	.42		

Note. For increase in R^2, $F = .36$, $F_{c.v.} \approx 3.00$.
[a]Adjusted R^2 were used. $R^2 = .27$.
$p < .001$.

Table 24

Results: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Level (Global, Local) Subcales Regressed on FLORIDA WRITES!

Variable	Regression Coefficient	Standard Error	Standardized Regression Coefficient
SPM[a]	.00	.01	-.04
ITBS[b]	.02	.00	.52*
Global	.00	.01	-.01
Local	.01	.01	.12

[a]SPM = Standard Progressive Matrices. [b]ITBS = Iowa Test of Basic Skills.
$p < .001$.

Table 25

Summary ANOVA Table for Full Model: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Leaning (Liberal, Conservative) Subscales Regressed on FLORIDA WRITES!

Source of Variance	df	SS	MS	F	R^{2a}
Model	4	21.82	5.46	13.21	.26*
Error	133	54.93	.41		

Note. For increase in R^2, $F = 1.34$, $F_{c.v.} \approx 3.00$.
[a]Adjusted R^2 were used. $R^2 = .28$.
$p < .001$.

Table 26

Results: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Leaning (Liberal, Conservative) Subcales Regressed on FLORIDA WRITES!

Variable	Regression Coefficient	Standard Error	Standardized Regression Coefficient
SPM[a]	- .01	.01	-.06
ITBS[b]	.02	.00	.55*
Liberal	.01	.01	.10
Conservative	.00	.01	-.01

[a]SPM = Standard Progressive Matrices. [b]ITBS = Iowa Test of Basic Skills.
$p < .001$.

Table 27

Summary ANOVA Table for Full Model: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Form (Hierarchic, Monarchic, Oligarchic, Anarchic) Subscales Regressed on FLORIDA WRITES!

Source of Variance	df	SS	MS	F	R^{2a}
Model	6	23.50	3.92	9.53	.27*
Error	132	54.24	.41		

Note. For increase in R^2, $F = 1.03$, $F_{c.v.} \approx 2.37$.
[a]Adjusted R^2 were used. $R^2 = .30$.
$p < .001$.

Table 28

Results: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Form (Hierarchic, Monarchic, Oligarchic, Anarchic) Regressed on FLORIDA WRITES!

Variable	Regression Coefficient	Standard Error	Standardized Regression Coefficient
SPM[a]	- .01	.01	-.06
ITBS[b]	.02	.00	.55*
Hierarchic	.00	.01	.01
Monarchic	.00	.01	.03
Oligarchic	.01	.01	.10
Anarchic	.01	.01	.13

[a]SPM = Standard Progressive Matrices. [b]ITBS = Iowa Test of Basic Skills.
$p < .001$.

Table 29

Summary ANOVA Table for Full Model: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Scope (Internal, External) Subscales Regressed on FLORIDA WRITES!

Source of Variance	df	SS	MS	F	R^{2a}
Model	4	22.38	5.60	13.55	.27*
Error	134	55.35	.41		

Note. For increase in R^2, $F = 1.72$, $F_{c.v.} \approx 3.00$.
[a]Adjusted R^2 were used. $R^2 = .29$.
$p < .001$.

Table 30

Results: Cognitive Ability (SPM), Iowa Test of Basic Skills, and Scope (Internal, External) Regressed on FLORIDA WRITES!

Variable	Regression Coefficient	Standard Error	Standardized Regression Coefficient
SPM[a]	- .01	.01	-.06
ITBS[b]	.02	.00	.52*
Internal	.01	.01	.09
External	.01	.01	.16

[a]SPM = Standard Progressive Matrices. [b]ITBS = Iowa Test of Basic Skills.
$p < .001$.

As the data in Tables 21 through 30 demonstrate, the addition of the TSI subscales to the SPM and ITBS did not increase prediction of students' FLORIDA WRITES! scores. ITBS was the only variable to explain a significant proportion of the variance in FLORIDA WRITES! scores. Cognitive style did not impact students' achievement scores when controlling for general cognitive ability.

To examine if the model containing both ITBS and SPM as predictors increased the R^2 over the model containing just SPM, a third set of hierarchical analyses was conducted. The reduced model contained just SPM as the predictor. The full model included both SPM and ITBS. The dependent variable was FLORIDA WRITES!. Tables 31 through 34 present results of these analyses.

Table 31

Summary ANOVA Table for Reduced Model: Cognitive Ability (SPM) Regressed on FLORIDA WRITES!

Source of Variance	df	SS	MS	F	R^{2a}
Model	1	4.37	4.37	8.25	.05*
Error	139	73.62	.53		

[a]Adjusted R^2 were used. $R^2 = .06$.
$p < .01$.

Table 32

Results of Reduced Model: Cognitive Ability (SPM) Regressed on FLORIDA WRITES!

	Regression	Standard	Standardized
Variable	Coefficient	Error	Regression Coefficient
SPM[a]	.02	.01	.23*

[a]SPM = Standard Progressive Matrices.
$p < .01$.

Table 33

Summary ANOVA Table for Full Model: Cognitive Ability (SPM) and Iowa Test of Basic Skills Regressed on FLORIDA WRITES!

Source of Variance	df	SS	MS	F	R^{2a}
Model	2	20.15	10.07	23.96	.25*
Error	137	57.59	.42		

Note. For increase in R^2, $F = 35.98$, $F_{c.v.} \approx 3.84$.
[a]Adjusted R^2 were used. $R^2 = .26$.
$p < .001$.

Table 34

Results of Full Model: Cognitive Ability (SPM) and Iowa Test of Basic Skills Regressed on FLORIDA WRITES!

Variable	Regression Coefficient	Standard Error	Standardized Regression Coefficient
SPM[a]	.00	.01	-.03
ITBS[b]	.02	.00	.52*

[a]SPM = Standard Progressive Matrices. [b]ITBS = Iowa Test of Basic Skills.
$p < .001$.

Results demonstrate that the addition of ITBS as a predictor variable did increase the amount of variance in FLORIDA WRITES! explained over SPM alone.

CHAPTER 4
DISCUSSION

Summary of Findings

Overall, the findings of this study did not support Sternberg's (1997b) contention that students' thinking styles interact with their performance on achievement tests. Sternberg (1990) asserted that "styles of thinking and learning are every bit as important as levels of ability..."(p. 367), and that "the teaching methods, the test formats [emphasis added], and the assignments cater to different styles (p. 367)." This suggests that awareness of students' thinking styles will help to explain their performance on achievement tests. It also suggests that style is as important as general cognitive ability in terms of the effect that both have on student achievement. The current study tested this hypothesis, using two different types of achievement tests - - a standardized, multiple-choice test of language skills (the ITBS) and a performance-based assessment of writing skills (the FLORIDA WRITES!). General cognitive ability was measured with the SPM, and cognitive style was assessed using Sternberg's TSI. It was found that when both general cognitive ability and thinking style were used to predict scores on the Iowa Test of Basic Skills (ITBS) and the FLORIDA WRITES! tests, only general cognitive ability was found to have a significant relationship with either achievement test. SPM scores correlated .24 with FLORIDA WRITES!, and .51 with ITBS scores, both significant at the .01 level. These correlations are consistent with previous research on the relationship

between SPM and achievement test scores (Sattler, 1992). Multiple regression analyses demonstrated that no domain of the TSI was able to significantly add to the variance in achievement test scores accounted for by general cognitive ability.

The lack of statistically significant findings from the regression analyses calls into question the validity of the instrument used to measure thinking style, the Thinking Style Inventory (TSI; Sternberg & Wagner, 1991), or the theory upon which it is based. The psychometric properties of the TSI have not been well-established, although the instrument has been used in several studies conducted by its developer (e.g., Grigorenko & Sternberg, 197; Sternberg & Grigorenko, 1995), and even in one cross-cultural study (Zhang & Sachs, 1997). The results of the present study suggested that an additional exploratory factor analysis of the TSI was warranted in an attempt to better understand what this instrument measures.

Factor Analysis

Purpose

The purpose of the factor analysis was to identify the constructs underlying the TSI that best explain the relationship between the subscale scores. Sternberg (1997b) based the TSI on his theory of mental self-government, and categorized each of the subscales accordingly. For example, in his theory Sternberg suggests that people's minds function like one of the three branches of the federal government: the legislative, executive, or judicial. Therefore, for the TSI he developed three subscales (Legislative, Executive, and Judicial) that are subsumed under the domain of Function. The theory of self-government only describes relationships between subscales subsumed under the same

domain, not relationships between subscales falling under different domains. For example, the theory suggests that under the Function domain, a person will have a preference for either the Legislative, Executive, or Judicial style. Sternberg does not suggest that certain styles "go together" - that is, he does not suggest, for example, that persons with an Executive style (categorized under the Function domain) also tend to be Conservative (categorized under the Leaning domain). In other words, he does not discuss style "profiles" that predict certain relationships between subscales of different domains. Therefore, when using a factor analysis to examine the congruence between Sternberg's theory of mental self-government and the instrument upon which it is based (the TSI), the primary relationships of interest should be those between subscale scores within each domain, not between subscales of different domains. For example, because Sternberg's theory suggests that people have a preference for either the Legislative, Executive, or Judicial style, a high positive correlation between scores on these subscales would not be expected. Therefore, in a factor analysis, the factors should demonstrate contrasts between these three subscales. The subscales should either load on different factors altogether, or, if they load strongly on the same factor, they should have contrasting positive and negative loadings. Similar contrasts should exist between the subscales within each domain.

A secondary purpose of the factor analysis was to see if perhaps certain style profiles do exist. That is, although Sternberg (1997b) does not make predictions regarding relationships beween subscales of differing domains, they may be present. In fact, separate analyses by Sternberg (1997b) and Zhang and Sachs (1997) both found relationships between subscales that were not predicted by the theory of mental self-government. An

additional analysis performed on scores from a larger and younger population than that used by the previous two studies may help elucidate these relationships.

Results of Factor Analysis

To examine the factor structure of the TSI, a principal-components factor analysis was performed on the intercorrelations of the 13 TSI subscales, using the SPSS statistical program. The correlation matrix of the subscales is found in Table 35. The vast majority of the subscale intercorrelations were positive and statistically significant, using an alpha level of .05. This suggests that there is little differentiation between the 13 subscales of the TSI.

Using Kaiser's (1958) mineigen criterion (eigenvalues greater than or equal to 1.0), three factors were initially considered for rotation ($\lambda_1 = 5.09$, $\lambda_2 = 1.9$, $\lambda_3 = 1.18$), accounting for 63% of the total variance in subscale scores (39%, 15%, and 9%, respectively). A Cattell's scree plot for the factor analysis of the 13 subscales was also generated. The scree plot was ambiguous, suggesting either a 2 or 3-factor solution. The Factor 1 loadings for the unrotated factor solution are in Table 36. The factor loadings represent the correlation coefficient between each subscale and the unrotated Factor 1. The factor loadings range from .28 (Internal subscale) to .76 (Judicial and Local subscales).

Table 35

Interscale Pearson Correlation Matrix for 13 Subscales of the Thinking Styles Inventory (N=153)

Subscale	1	2	3	4	5	6	7
1. Legislative							
2. Executive	.12						
3. Judicial	.36**	.47**					
4. Global	.39**	.21*	.32**				
5. Local	.36**	.46**	.52**	.34**			
6. Liberal	.51**	.22**	.56**	.42**	.51**		
7. Conservative	.07	.69**	.38**	.27**	.36**	.08	
8. Hierarchic	.31**	.47**	.50**	.32**	.53**	.45**	.31**
9. Monarchic	.21**	.36**	.21**	.30**	.42**	.21**	.37**

*p < .05. **p < .01.

Table 35 - continued

Subscale	8	9	10	11	12	13
9. Monarchic	.43**					
10. Oligarchic	.34**	.35**				
11. Anarchic	.49**	.33**	.31**			
12. Internal	.26**	.27**	-.14	.20*		
13. External	.30**	.22**	.45**	.48**	-.19*	

*p < .05. **p < .01.

Table 36

Factor 1 Loadings for the Unrotated Factor Solution

Subscale	Factor 1 Loading
Legislative	.52
Executive	.67
Judicial	.76
Global	.55
Local	.76
Liberal	.68
Conservative	.56
Hierarchic	.72
Monarchic	.57
Oligarchic	.60
Anarchic	.71
Internal	.28
External	.59

After examining the eigenvalues, scree plot, and the factor loadings and communalities for the 2 and 3-factor solutions, the 2-factor model was determined to be the best fit of the data. This was primarily determined by the increased interpretability of the 2-factor over the 3-factor model.

The factor loadings, communalities, and uniqueness estimates for the 2- and 3-factor solution are found in Tables 37 and 38, respectively.

Table 37

Varimax-Rotated Two-Factor Model for Thinking Styles Inventory

Scale	Factor 1	Factor 2	h^2	u^2
Legislative		.74	.55	.45
Executive	.81		.67	.33
Judicial	.57	.51	.59	.41
Global		.57	.32	.68
Local	.53	.54	.57	.43
Liberal		.78	.62	.38
Conservative	.78		.61	.39
Hierarchic	.49	.54	.53	.47
Monarchic	.44	.36	.32	.68
Oligarchic	.80		.64	.36
Anarchic	.43	.59	.53	.47
Internal		.70	.49	.51
External	.60		.41	.59
% of Variance	27.6	23.6		
Cumulative %	27.6	51.2		
Eigenvalue	5.1	1.9		

Note. Factor loadings of less than |.30| have been omitted.

The factor loadings represent correlation coefficients between the subscales and each factor. For example, the correlation between the legislative subscale and Factor 2 is .74. Only factor loadings higher than |.30| are included in the table. The communalities (h^2) demonstrate how much of the variance in the subscales is accounted for by the extracted factors. To illustrate, the 2-factor model accounts for 55% of the variance in the Legislative subscale. For the 2-factor model, the communalities ranged from .32 (Global and Monarchic subscales) to .67 (Executive subscale). They ranged from .38 (Global) to .76 (Internal) for the 3-factor model. This suggests that both models are only moderately successful in accounting for the subscale variance.

The uniqueness (u^2) represents the portion of each subscales' variance that is not accounted for by the factor model. Uniqueness is equal to 1 - h^2. Thus, for the 2-factor model, the u^2 for Legislative subscale calculated by using the following formula: 1.0 - .55 = .45. This means that in the 2-factor model, 45% of the variance in the Legislative subscale is not accounted for by either Factor 1 or Factor 2. This unique variance may be a portion of the subscale's true score variance that is unrelated to any other subtest (specific variance), or it may be due to error (error variance). In a test with low reliability, much of the unique variance will be due to error. For the 2-factor model, the uniqueness estimates ranged from .68 (Global and Monarchic subscales) to .33 (Executive subscale). Thus, the 2-factor model does not account well for the variance in the Global and Monarchic subscales, but it accounts reasonably well for the variance in the Executive subscale. The 3-factor model is only slightly better than the 2-factor model in accounting for subscale variance. In the 3-factor model, the uniqueness ranges from .62 (Global) to .24 (Internal).

Thus, of the 13 subscales, it appears that the Global subscale is the least explained by the factor analysis.

Table 38

Varimax-Rotated Three-Factor Model for Thinking Styles Inventory

Scale	Factor 1	Factor 2	Factor 3	h^2	u^2
Legislative	.68	.32		.56	.44
Executive		.82		.72	.28
Judicial	.69	.38		.64	.36
Global	.51			.38	.62
Local	.55	.51		.59	.41
Liberal	.85			.73	.27
Conservative		.84		.71	.29
Hierarchic	.51	.49		.55	.45
Monarchic		.61	.42	.59	.41
Oligarchic		.68	-.32	.64	.36
Anarchic	.72			.59	.41
Internal	.31		.81	.76	.24
External	.61		-.53	.72	.28
% of Variance	27.6	23.6	11.7		
Cumulative %	27.6	51.2	62.8		
Eigenvalue	5.1	1.9	1.2		

Explanation of Subscales by Factor Models

The 2-factor model is consistent with some aspects of Sternberg's theory. In particular, the model demonstrates that many of the subscales contrast in the expected manner. For example, under the Function domain, the two factors clearly differentiate between the subscales of Executive and Legislative. The Executive subscale has a factor loading of .81 on Factor 1, whereas the Legislative subscale has a factor loading of .74 on Factor 2. The Judicial subscale is not explained by either Factor 1 or 2 (or Factor 3 in the 3-factor model).

Under the domain of Level, the Global subscale is predominantly related to Factor 2. The Local subscale, however, loads equally on both factors (.53 on Factor 1; .54 on Factor 2). For the Leaning domain, there is a clear differentiation between the Liberal and Conservative subscales. Factor 1 has a clear relationship with the Conservative subscale (.78), while the Liberal subscale loads as strongly on Factor 2 (.78).

Under the domain of Form, the only subscale that loads strongly on a single factor is the Oligarchic subscale (.80 loading on Factor 1). The other three subscales load equally well on both factors. Finally, the 2-factor model demonstrates a clear differentiation between the Internal and External subscales of the Scope domain. Factor 1 is dominated by the External subscale (.60), whereas Factor 2 is more strongly related to the Internal subscale (.49).

Student Profiles Constructed from Factor Models

By examining the dominant loadings on each factor, an interpretation of the 2-factor model is possible. Factor 1 is dominated by the subscales of Executive,

Conservative, Oligarchic, and External (factor loadings of .81, .78, .80, and .60, respectively). This suggests that for this sample, a style profile exists that characterizes a type of student that prefers to follow rules, likes problems that are prestructured or prefabricated, and prefers to fill in gaps within existing structures rather than creating structure themselves. They like to do things in a familiar, traditional way rather than making up their own rules or procedures. Students fitting this profile also prefer to do many things at once, and often have difficulty prioritizing. Lastly, they tend to be more extroverted than introverted.

The second student profile is, in many ways, a direct contrast to the first profile. This profile is dominated by those students with a Legislative, Global, Liberal, and Internal style (factor loadings of .74, .57, .78, .79, respectively). Students with this profile like to create and formulate things, and they tend to use their own methods of problem-solving, rather than relying on prestructured or prefabricated solutions. They also prefer working with abstract concepts and ideas more than details. Socially, they are more introverted than their counterparts, preferring to work alone and be more task than people-oriented.

The 3-factor model can be interpreted in the same way as the 2-factor model, but as stated earlier, it does not yield as "clean" an interpretation. In terms of profiles, Factor 2 of the 3-factor model appears very similar to Factor 1 of the 2-factor model. For example, in the latter model Factor 2 is dominated by the Executive, Conservative, and Oligarchic subscales (factor loadings of .82, .84, and .68, respectively). Unlike the 2-factor model, however, the External subscale is not included in this profile (factor loading less than .30). Thus this profile characterizes a type of student that prefers to follow rules, and

likes solving problems that are prestructured or prefabricated. They prefer using familiar, traditional methods rather than making up their own rules or procedures. Students fitting this profile also prefer to do many things at once, and often have difficulty prioritizing.

The second profile, taken from Factor 1 of the 3-Factor model, is represented by students with Legislative, Judicial, Liberal, and Anarchic styles (factor loadings of .68, .69, .85, and .72, respectively). One difficulty in interpreting this profile is the lack of contrast between the Legislative and Judicial styles. Both fall under the domain of Function, and according to the theory of mental self-government, they should either load on different factors, or should have opposite positive and negative factor loadings. Instead, they have practically equal factor loadings (.68 for Legislative, .69 for Judicial). This profile is similar to its 2-factor counterpart, in that both contain the Legislative and Liberal styles. However, whereas the 2-factor profile also contains the Global and Internal styles, the 3-factor profile is dominated by the Anarchic. Thus, students fitting this second style profile like creative tasks and using their own methods to problem-solve. They also like to evaluate rules and procedures, and to analyze situations. In terms of approaching problems, they take a random approach, and do not like being restricted by systems or rules.

The third factor is dominated by the Internal subscale (factor loading of .81), and, to a lesser degree (i.e., smaller factor loadings), it also represents contrasts between the Internal and External (-.53) subscales and the Monarchic (.42) and Oligarchic (-.32) subscales. Thus, this profile represents two types of students: one who is Monarchic and Internal and one who is Oligarchic and External. The first characterizes a student who is

introverted and prefers to focus on a single project to the exclusion of others. The opposite would be the student who is extroverted and prefers to focus on many projects at once.

Summary of Factor Analysis Findings

An exploratory analysis of the factor structure of the TSI did not support the 5-domain structure (viz., Function, Form, Level, Scope, and Leaning). Rather, a 2-factor model was the best fit for the data, although it was unable to account for 49% of the variance in the subscale scores. Factor 1 had strong associations with the subscales of Executive, Conservative, Oligarchic, and External. Factor 2 is best represented by the subscales of Legislative, Global, Liberal, and Internal. Thus, for this sample, the TSI yields two student "profiles": the first profile characterizes students who prefer to follow rules, like problems that are prestructured, and prefer to fill in gaps within existing structures rather than creating structure themselves. These students like to do things in familiar, traditional ways rather than making up their own rules or procedures. They prefer to do many things at once, rather than prioritizing, and then tend to be more extroverted than introverted.

The second profile describes students who like to create and formulate things, and tend to use their own methods of problem-solving rather than rely on traditional solutions. They prefer working with abstract concepts and ideas more than details. In addition, they tend to be more introverted than extroverted.

Sternberg's theory of self-government (1997b) does not specifically describe predictable "patterns" or conglomerates of styles (for example, that Executive persons

also tend to be External), so the pattern of factor loadings found in the current study cannot be compared to his theory. In his own analysis, however, Sternberg found a 5-factor model with some similarities to the 2-factor model found in the present study. For example, Sternberg's Factor 1, Adherence to Structure, contrasted the Liberal and Legislative scales with the Conservative and Executive. He concluded that "liberal people tend, on average, to be legislative, whereas conservative people tend to be executive" (p. 126). The present study showed a similar pattern. Executive and Conservative were related, but also included in this relationship were Oligarchic and External styles. In the same way, Legislative and Liberal were found to have a strong relationship, but also included were Global and Internal styles. Thus, the current study found profiles that were more inclusive than than those reported by Sternberg.

Sternberg's second factor, Engagement, comprised the two scales of Oligarchic (inverted) and Judicial. This finding was not replicated in the current study. Factor III, Scope, contrasted the External and Internal sclaes. The present study did not find a single factor that contrasted these two scales. Rather, the subscales loaded on separate factors and thus were subsumed under different profiles. Factor IV, Level, contrasted the Local and Global scales. In the present study, the Global scale was predominantly related to Factor 2; the Local subscale, however, loaded equally well on both factors, thus not showing a clear contrast between the two subscales. Lastly, Sternberg's fifth factor, Distribution of Time, comprised the Hierarchical subscale. The Hierarchical subscale did not show differentiated factor loadings in the present study.

Although the theory of mental self-government does not elaborate on student types, it does state that within each domain, the subscales should contrast one another. For example, under the domain of Leaning, persons who are Liberal should not also be Conservative - the two are (theoretically) diametrically opposed. Through examination of the factor structure, it was found that many of the domains did show the expected contrasts between the subscales. In particular, the 2-factor model differentiated between the subscales of the Leaning (Liberal vs. Conservative) and Scope (Internal vs. External) domains. Within the Function domain, the Executive and Legislative subscales clearly loaded on different factors; within that same domain, however, the Judicial subscale was not well-accounted for by the model. The domains of Level (Global vs. Local) and Form (Hierarchic, Monarchic, Oligarchic, and Anarchic) were not supported by the factor analysis.

Results of this exploratory factor analysis were very similar to those of Zhang and Sachs (1997). They reported a 3-factor solution that accounted for 66% of the variance in subscale scores. Like the current study, they also found that the domains of Form and Level were not clearly identified by the model. An examination of the correlation matrix of the TSI subscale scores also showed more similarities between the findings of the current study and those of Zhang and Sachs (1997) than those of Sternberg (1997b). For example, Sternberg reported negative correlations between the subscales of Global and Local, and also between the Legislative and Conservative subscales. Neither of these findings has been replicated. The studies of Zhang and Sachs and that of Sternberg both demonstrated a negative correlation between the Liberal and Conservative subscales, whereas the

current study did not. All three studies, however, found significant ($p < .01$), positive correlations between the subscales of Liberal and Legislative, and between Conservative and Executive. In sum, portions of Sternberg's theory have evidence of construct validity (based on results of the current study and that of Zhang and Sachs), where others clearly do not.

The communalities of the 13 subscales were low to moderate, with the largest being .67 (Executive). This suggests that a large proportion of variance in subscale scores is due to either specific variance (the portion of total variance that is unrelated to the other variables) or error variance. In other words, even the best factor structure found (the 2-factor model) did not find strong clusters among the subscale inter-correlations. This could be due to either having subscales that are not related to one another in predictable ways (high uniqueness), or in having subscales with low reliabilities (high error).

Subscale Reliabilities

To examine the reliability of the TSI, Cronbach's alpha estimates were computed for each of the 13 subscales. The median was .65, suggesting that the items within the subscales are not homogeneous (i.e., they are not measuring the same thing). As an example, the Global subscale ($\alpha = .51$) consists of the following five items: (item #7) "I tend to pay little attention to details;"(item #18) "I care more about the general effect than about the details of a task I have to do;" (item #38) "I tend to emphasize the general aspect of issues or the overall effect of a project;" (item #48) "I like working on projects that deal with general issues and not with nitty-gritty details;" and (item #61) "In talking or writing down ideas, I like to show the scope and context of my ideas, that is, the

general picture." Item 7 has the lowest correlation with the subscale total score (r = .03). Analyses show that if that item were dropped, the alpha would increase to .64. Descriptive analyses demonstrate that the mean ratings for items 18, 38, 48, and 61 range from 4.0 to 4.5 ("somewhat well" to "well"), whereas the mean rating for item 7 is 3.0 ("slightly well"). Several possible explanations exist for why item 7 is inconsistent with the others within the Global subscale: one, item 7 is shorter in length than the others, and therefore offers more clarity to the reader. Also, the other four items all contain the word "general", whereas item 7 focuses on "details". Similar problems were found within the other TSI subscales.

Implications for Research on the Construct of Style

The weak psychometric properties of the TSI confound the original question of whether thinking styles influence students' performance on different types of tests. The instrument used in this study and the theory upon which it is based are both relatively new and have been subject of very few independent studies. The current study, in part, served as a study of the validity of Sternberg's theory of thinking styles and the TSI instrument. Because of its questionable validity, the use of the TSI as an independent measure casts doubts on the finding that thinking style and test type are unrelated. In other words, a more psychometrically-sound instrument may have yielded different results. Unfortunately, although there are literally dozens of measures of thinking style, none appear to have adequate empirical support (see DeBello, 1989, & Sewall, 1986, for reviews of instruments).

The fundamental difficulty in interpreting the results of this study is in trying to escape the circular logic found at the center of the issue. Again, there are no measures of thinking style that have adequate empirical support. In the decades following the very first measure of thinking style, researchers have attempted to gather evidence of validity by performing studies that are fundamentally very similar to the current one. That is, they attempted to use their instrument to predict some aspect of participant performance. For example, research has attempted to use learning style measures to predict student performance in different learning environments (e.g., environments characterized by different teaching styles, educational materials, sensory modalities). When their study fails to demonstrate an interaction between style and student performance, the researcher is left with the same question as that raised by the current study: is the instrument at fault, or is there truly no relationship between student styles and academic performance? When looking at each study individually, it is prudent (and tempting) to conclude that the instrument is at fault - that the instrument is failing to tap into the underlying construct of style. However, when the entire body of research in this area is collectively examined, a much different picture emerges. Style is simply not substantiated as a viable and important construct. Although it may have face validity (as evidenced by the continued interest of both researchers and practitioners), the construct of style has failed in numerous studies, including the present one, to mediate student performance (for a meta-analysis, see Kavale & Forness, 1987).

Because it is so intuitively appealing, style will probably persevere as a construct considered by many as important to effective teaching and learning. Researchers of this

topic may wish to refine the theory of style. For example, most of the studies thus far imply that students' preferences (i.e., style) will influence their academic performance. However, it is possible that simply preferring one type of instruction or assessment mode over another does not necessarily translate into improved learning or performance. For example, in the context of Sternberg's (1997b) theory, perhaps students do differ on the constructs of Executive and Legislative style: it is plausible that some students prefer following rules and solving prefabricated problems (Executive), while others prefer creating and formulating ideas (Legislative). But a student's preference for one type of activity over another does not necessarily translate into improved performance on the preferred over the non-preferred activity. In other words, a student may like multiple-choice tests more than essay tests, but he actually performs better on multiple-choice test. An examination of the wording of the TSI items demonstrates that it clearly was developed to measure students' preferences, not what students perceive to be their strengths: 42 out of the 65 items on the TSI use the words "prefer," "like," "enjoy," or "care." None of the items use the words "able," "better," or "worse." This choice in wording is assumed to be intentional, as Sternberg (1997b) describes thinking styles as preferences, not abilities. Therefore, it is possible that the construct of thinking style does exist, and that the TSI measures it adequately. However, the assumption that accomodating students' thinking style will somehow lead to improved academic performance is unsubstantiated.

If, in fact, style as a construct exists but is not related to academic performance (a question that could be answered through additional research), then perhaps its role in

education needs to be re-examined. Educators and educational researchers are generally interested in identifying those variables that help students to succeed, and success in the system is often synonymous with high academic performance. Schools and school systems are judged and evaluated by the performance of their students on academic achievement tests, and students are frequently evaluated, promoted, retained, and rewarded on the basis of their academic achievement (Resnick & Resnick, 1991). Student interests and preferences are of minimal interest to most researchers unless they are demonstrated to affect academic achievement (i.e., test scores and grades). Therefore, if thinking style is not substantiated as a variable affecting academic achievement, then its usefulness to educators may be minimal.

Implications for Future Research on Performance Assessments

Results of this study have implications for future studies of performance assessments. Performance assessment is a type of educational measurement that has received increased interest and popularity over the past decade (Council of Chief State School Officers, 1998). Lauded for their authenticity and emphasis on higher-order thinking skills, performance assessments are now replacing or supplanting traditional, multiple-choice standardized achievement tests in the majority of the state educational testing programs (Council of Chief State School Officers, 1998). Most of the research thus far has examined the psychometric properties of specific performance assessment instruments (e.g., Koretz et al., 1994; LeMahieu et. al, 1995). Fewer studies have examined the student characteristics that predict performance on these types of tests. It makes intuitive sense that some students will prefer the format of performance

assessments over that of traditional tests (or vice versa). Typically, performance assessments offer less structure and more ambiguity than multiple-choice tests, affording students the opportunity to construct responses rather than select answers (Resnick & Resnick, 1991; Wiggins, 1989). This increased freedom in response may appeal to some students, but may intimidate others. The preferences of students regarding test format (i.e., performance vs. multiple-choice test) could easily be measured by a well-developed self-report instrument. What remains to be proven is whether matching preference leads to improved performance. The present study does not support such a link, but as mentioned above, the instrument used is of questionable validity. However, if the popularity of performance assessments continues, further studies are needed on the variables that predict poor and exceptional performance, whether that variable is style or some other aspect of the student or the educational environment.

Limitations of the Current Study

The primary limitation of this study is its lack of tenable findings. That is, the lack of significant results regarding the interaction between thinking style and performance on differing test types leads to one of two conclusions: either the TSI was an inappropriate (i.e., invalid) instrument for this study, or that there is, in fact, no substantial relationship between thinking style and academic performance. Future research could elucidate this issue by developing a measure of thinking style that meets generally accepted standards of psychometric properties (viz., reliability and validity). Such an instrument would, of course, confront the same issue as its predecessors regarding the expectation that it should predict performance on the criterion variable. In other words, researchers will judge the

instrument by its ability to differentially predict student performance. If it fails, then its validity will most likely be questioned. To satisfactorily conclude that style is a viable and important construct, an instrument must be developed that demonstrates an interaction between assessed style and student performance. An experimental study should demonstrate that student performance is improved by considering and accomodating style. Thus far, no study has succeeded in this.

Other weaknesses of the current study include its limited sample, the dependent measures used, and the type of statistical analyses employed. The sample was restricted to approximately 150 students drawn from a small county in the southeastern region of the United States. In addition, the population was fairly homogeneous in terms of race. Generalizability of the findings to students from other parts of the country or to larger school districts would be problematic, as would assuming that the findings would hold true for a more heterogeneous population. This is primarily a question of external validity.

The dependent measures (ITBS and FLORIDA WRITES!) in this study were chosen because they appeared to measure the same skill through different means. The language portion of the ITBS reportedly measures the basic skills common to standard written English (ITBS Technical Manual, 1994). This includes spelling, capitalization, punctuation, word usage, and expression. The format of the test is multiple-choice, and is scored mechanically using a correct/incorrect dichotomy. The FLORIDA WRITES! test also assesses students' mastery of writing skills, but does so by asking students to compose a reponse to a writing prompt. FLORIDA WRITES! items are judged by raters trained in the use of a scoring rubric that emphasizes focus, organization, support for

ideas, and the conventions of written English. No data were available prior to this study regarding the relationship between scores on FLORIDA WRITES! and standardized, multiple-choice measures of writing skill. The current study found that the relationship between the two measures (for this sample) was .50 ($p < .01$), indicating approximately 25% shared variance between the instruments. This finding may be viewed in multiple ways. One, the significant correlation between the two supports the idea that the tests are both assessing a similar construct, presumably language skills. This finding was expected. In addition, it is also a positive sign that the two did not have a near-perfect (i.e., approaching 1.0) correlation, because that would have implied redundancy, which would not bode well for the state that has invested substantial finances into administering two separate achievement tests. The current study, however, rested on the assumption that the two tests were measuring the same skill, but through different means. Given that some students may perform significantly better on one test than another, the hypothesis was that one variable contributing to the intraindividual variation may be style, or a preference for one test format over the other. If style had been substantiated as a mediating variable, then the implication would be that students' preferences for testing format affect their scores. The moderate correlation between the two, however, raises the question of just how much similarity there is between the skills measured on the ITBS and the FLORIDA WRITES!. It may be that they are measuring different constructs altogether, and that the shared variance is due to some other variable. Results of this study suggest that the other variable may be something other than general cognitive ability. When scores on the SPM were added into the regression equation, the partial correlation coefficient between the ITBS

and FLORIDA WRITES! remained significant ($B = .52$, $p < .01$). Further research could be conducted on the relationship between FLORIDA WRITES! and other measures of language skill.

Finally, the statistical analyses used may be considered a limitation of the study. To address the factor structure of the TSI, a confirmatory factor analysis (CFA) may have been preferable over an exploratory factor analysis (EFA). Bandalos (1996) summarized the different uses of CFA and EFA, and said that EFA is appropriate when (a) the literature base of an instrument is weak, (b) the number of factors needs to be determined, (c) the factor loadings are of interest, and (d) the researcher wants the items to load freely on all factors. CFA, on the other hand, is acceptable when (a) the theoretical base is strong, (b) the number of factors is fixed, (c) the degree of factor correlations is fixed a priori, and (d) variables will be fixed to load onto specific factors(s). EFA was chosen for this study primarily because of the weak theoretical and literature base of the TSI. Only two published studies report results of factor analyses for the TSI (Sternberg, 1997b; Zhang & Sachs, 1997). Both were EFA's performed at the subscale level. Sternberg's (1997b) study yielded a five-factor model in which three of the factors were consistent with his theory. In contrast, Zhang and Sachs (1997) found that a three-factor model was the best fit for the data. In that study, some of the factor loadings were in the direction predicted by Sternberg (1997b), whereas others were inconsistent with theory. Taken together, the results of these two studies yield an unclear pictue of the factor structure of the TSI, therefore warranting the EFA conducted in the present study. Results of the present study were more consistent with those of Zhang and Sachs (1997), not

Sternberg (1997b). With a larger sample size, both an EFA and CFA could have been used by splitting the data in half. The initial EFA would discover a feasible factor structure. A CFA conducted with the second half of the data could then "simplify, refine, and confirm" the model (Bryant & Yarnold, 1997, p. 109). In future research, this type of dual analysis would be helpful in clarifying the underlying factor structure of the TSI.

APPENDIX A
CORRELATIONS OF ABILITIES, ACADEMIC PERFORMANCE, AND STYLES

	Performance					
Item	Analytic	Creative	Practical	Homework	Exams	Project
Abilities						
[a]STAT-Analytic	.25*	.34*	.34*	.26*	.27*	.31*
STAT-Creative	.15*	.21*	.20*	.13	.16*	.20*
STAT-Practical	.15*	.17*	.17*	.12	.18*	.15*
Thinking Styles						
Legislative	.17*	.16*	.14	.12	.14*	.17*
Judicial	.15*	.20*	.23*	.21*	.18*	.15*
Executive	-.15*	-.16*	-.10	-.12	-.07	-.18*
Monarchic	-.06	-.06	-.08	-.05	-.04	-.10
Hierarchic	.06	.16*	.11	.13	.07	.08
Oligarchic	-.06	-.11	-.11	-.12	-.11	.03
Anarchic	-.08	-.13	-.12	-.10	-.08	-.12
Local-Global	.07	.04	.09	.05	.01	.12
Liberal-Conservative	.14	.10	.08	.03	.09	.16

Note: From "Styles of Thinking, Abilities, and Academic Performance," by Grigorenko & Sternberg, 1997, Exceptional Children, 63, p. 295.
[a]STAT = Sternberg Triarchic Abilities Test, Level H. * $p < .05$.

APPENDIX B
THINKING STYLES INVENTORY AND RESPONSE SHEET

Thinking Styles Inventory

This is a questionnaire about the different strategies and ways people use to solve problems, to carry out tasks or projects, and to make decisions. Read each statement carefully and decide how well it describes you. Use the scale provided (on the **Response Sheet**) to indicate how well the statement fits the way you typically do things at school, at home, or on a job. Circle 1 if the statement does **not** fit you at all, that is, you almost never do things this way. Circle 7 if the statement fits you extremely well, that is, you almost always do things this way. Use the values in between to indicate that the statement fits you in varying degrees.

There are, of course, no right or wrong answers. Please read each statement and circle the number on the scale on the response sheet that best indicates how well the statement describes you.

Please proceed at your own pace, but do not spend too much time on any one statement.

If you have any questions, feel free to ask them now.

1. I prefer to deal with specific problems, rather than with general questions.

2. When talking or writing about ideas, I stick to one main idea.

3. When starting a task, I like to brainstorm ideas with friends or peers.

4. I like to set priorities for the things I need to do before I start doing them.

5. When faced with a problem, I use my own ideas and strategies to solve it.

6. In discussing or writing on a topic, I think the details and facts are more important than the overall picture.

7. I tend to pay little attention to details.

8. I like to figure out how to solve a problem following certain rules.

9. I like to control all phases of a project, without having to consult with others.

10. I like to play with my ideas and see how far they can go.

11. I am careful to use the proper method to solve any problem.

12. I enjoy working on things that I can do by following directions.

13. I stick to standard rules or ways of doing things.

14. I like problems where I can try my own way of solving them.

15. When trying to make a decision, I rely on my own judgment of the situation.

16. I can switch from one task to another easily, because all tasks seem to me to be equally important.

17. In a discussion or report, I like to combing my own ideas with those of others.

18. I care more about the general effect than about the details of a task I have to do.

19. When working on a task, I can see how the parts relate to the overall goal of the task.

20. I like situations where I can compare and rate different ways of doing things.

21. When there are many important things to do, I try to do as many as I can in whatever time I have.

22. When I'm in charge of something, I like to follow methods and ideas used in the past.

23. I like to check and rate opposing points of view or conflicting ideas.

24. I like to collect detailed or specific information for projects on which I work.

25. In dealing with difficulties, I have a good sense of how important each of them is and in what order to tackle them.

26. I like situation where I can follow a set routine.

27. When discussing or writing about a topic, I stick to points of view accepted by my colleagues.

28. I like tasks and problems that have fixed rules to follow in order to complete them.

29. I prefer to work on a project or task that is acceptable to and approved by my peers.

30. When there are several important things to do, I do those most important to me and my colleagues.

31. I like projects that have a clear structure and a set plan and goal.

32. When working on a task, I like to start with my own ideas.

33. When there are many things to do, I have a clear sense of the order in which to do them.

34. I like to participate in activities where I can interact with others as part of a team.

35. I like to tackle all kinds of problems, even seemingly trivial ones.

36. When faced with a problem, I like to solve it in a traditional way.

37. I like to work alone on a task or a problem.

38. I tend to emphasize the general aspect of issues or the overall effect of a project.

39. I like to follow definite rules or directions when solving a problem or doing a task.

40. When discussing or writing down ideas, I use whatever comes to mind.

41. When working on a project, I like to share ideas and get input from other people.

42. I like projects where I can study and rate different views of ideas.

43. When trying to make a decision, I tend to see only one major factor.

44. I like problems where I need to pay attention to details.

45. I like to challenge old ideas or ways of doing things and to seek better ones.

46. I like situations where I interact with others and everyone works together.

47. I find that solving one problem usually leads to many other ones that are just as important.

48. I like working on projects that deal with general issues and not with nitty-gritty details.

49. I like situation where I can use my own ideas and ways of doing things.

50. If there are several important things to do, I do the one most important to me.

51. I prefer tasks or problems where I can grade the design or methods of others.

52. When there are several important things to do, I pick the ones most important to my friends and colleagues.

53. When faced with a problem, I prefer to try new strategies or methods to solve it.

54. I like to concentrate on one task at a time.

55. I like projects that I can complete independently.

56. When starting something, I like to make a list of things to do and to order the things by importance.

57. I enjoy work that involves analyzing, grading, or comparing things.

58. I like to do things in new ways not used by others in the past.

59. When I start a task or project, I focus on the parts most relevant to my peer group.

60. I have to finish one project before starting another one.

61. In talking or writing down ideas, I like to show the scope and context of my ideas, that is, the general picture.

62. I pay more attention to parts of a task than to its overall effect or significance.

63. I prefer situations where I can carry out my own ideas, without relying on others.

64. I like to change routines in order to improve the way tasks are done.

65. I like to take old problems and find new methods to solve them.

Thinking Styles Response Sheet

Name: _____ *School*:_____

	1 Not at All Well	2 Not Very Well	3 Slightly Well	4 Somewhat Well	5 Well	6 Very Well	7 Extremely Well
1.	1	2	3	4	5	6	7
2.	1	2	3	4	5	6	7
3.	1	2	3	4	5	6	7
4.	1	2	3	4	5	6	7
5.	1	2	3	4	5	6	7
6.	1	2	3	4	5	6	7
7.	1	2	3	4	5	6	7
8.	1	2	3	4	5	6	7
9.	1	2	3	4	5	6	7
10.	1	2	3	4	5	6	7
11.	1	2	3	4	5	6	7
12.	1	2	3	4	5	6	7
13.	1	2	3	4	5	6	7
14.	1	2	3	4	5	6	7
15.	1	2	3	4	5	6	7
16.	1	2	3	4	5	6	7
17.	1	2	3	4	5	6	7
18.	1	2	3	4	5	6	7
19.	1	2	3	4	5	6	7
20.	1	2	3	4	5	6	7
21.	1	2	3	4	5	6	7
22.	1	2	3	4	5	6	7
23.	1	2	3	4	5	6	7
24.	1	2	3	4	5	6	7
25.	1	2	3	4	5	6	7
26.	1	2	3	4	5	6	7
27.	1	2	3	4	5	6	7
28.	1	2	3	4	5	6	7
29.	1	2	3	4	5	6	7
30.	1	2	3	4	5	6	7
31.	1	2	3	4	5	6	7
32.	1	2	3	4	5	6	7
33.	1	2	3	4	5	6	7

	1 Not at All Well	2 Not Very Well	3 Slightly Well	4 Somewhat Well	5 Well	6 Very Well	7 Extremely Well
34.	1	2	3	4	5	6	7
35.	1	2	3	4	5	6	7
36.	1	2	3	4	5	6	7
37.	1	2	3	4	5	6	7
38.	1	2	3	4	5	6	7
39.	1	2	3	4	5	6	7
40.	1	2	3	4	5	6	7
41.	1	2	3	4	5	6	7
42.	1	2	3	4	5	6	7
43.	1	2	3	4	5	6	7
44.	1	2	3	4	5	6	7
45.	1	2	3	4	5	6	7
46.	1	2	3	4	5	6	7
47.	1	2	3	4	5	6	7
48.	1	2	3	4	5	6	7
49.	1	2	3	4	5	6	7
50.	1	2	3	4	5	6	7
51.	1	2	3	4	5	6	7
52.	1	2	3	4	5	6	7
53.	1	2	3	4	5	6	7
54.	1	2	3	4	5	6	7
55.	1	2	3	4	5	6	7
56.	1	2	3	4	5	6	7
57.	1	2	3	4	5	6	7
58.	1	2	3	4	5	6	7
59.	1	2	3	4	5	6	7
60.	1	2	3	4	5	6	7
61.	1	2	3	4	5	6	7
62.	1	2	3	4	5	6	7
63.	1	2	3	4	5	6	7
64.	1	2	3	4	5	6	7
65.	1	2	3	4	5	6	7

APPENDIX C
TESTING TIMES AND ITEM COUNTS FOR IOWA TEST OF BASIC SKILLS

Testing Times and Number of Items on the Iowa Test of Basic Skills, Form K, Complete Battery

Test	Testing Time in Minutes	Number of Items
Spelling	12	41
Capitalization	30	12
Punctuation	30	12
Usage and Expression	43	24

Source: In Riverside Publishing Company, 1994, Riverside 2000 Integrated Assessment Program Technical Summary I, p. 16.

APPENDIX D
SCORING RUBRIC FOR FLORIDA WRITES!

Focused Holistic Rubric: How FLORIDA WRITES! Responses are Scored

The students' responses are scored using the holistic method. In this method of scoring, trained readers evaluate a piece of writing for its overall quality. Holistic scoring requires readers to evaluate the work as a whole, while considering four elements: focus, organization, support, and conventions. In this type of scoring, readers are trained not to become overly concerned with any one aspect of writing but to look at the response as a whole.

Focus refers to how clearly the paper presents and maintains a clear idea, theme, or unifying point. Papers representing the higher end of the point scale demonstrate a consistent awareness of the topic and do not contain extraneous information. Papers presenting the middle and lower ends of the point scale contain either loosely related information or extraneous information.

Organization refers to the structure or plan of development (beginning, middle, and end) and whether the points logically related to one another. Organization refers to the use of transitional devices (terms, phrases, and variation in sentence structure) to (1) signal the relationship of the supporting ideas to the main idea, theme, or unifying point and (2) the evidence of the connection and movement between sentences. Papers

representing the higher end of the point scale use transitions to signal the plan or text structure and end with summary or concluding statements. Papers representing the middle and lower ends of the point scale may lack transitional devices or concluding statements.

Support refers to the quality of the details used to explain, clarify, or define. The quality of the support depends on word choice, specificity, depth, credibility, and thoroughness. Papers representing the higher end of the point scale provide fully developed elaborated examples and illustrations and the relationship between the supporting ideas and the topic is clear. Papers representing the middle and lower ends of the point scale may contain support that is a bare list of events or reasons or support extended by detail.

Conventions refer to the mechanics of punctuation, capitalization, spelling, and variation in sentence structures used in the paper. These conventions are basic writing skills included in Florida's Language Arts curriculum Frameworks. Papers representing the higher end of the point scale follow, with few exceptions, the conventions of punctuation, capitalization, and spelling and use a variety of sentence structures to present ideas. Papers representing the middle and lower ends of the point scale may contain some or man errors in punctuation, capitalization, spelling, and sentence structure and may have little variation in sentence structure.

Source: Florida Statewide Assessment Program, 1995. Florida Writes! 1995 Technical Report, p. 8.

APPENDIX E
EIGHTH-GRADE SCORING RUBRIC FOR FLORIDA WRITES!

Score of 1: The writing may only minimally address the topic. The paper is fragmentary or incoherent listing of related ideas or sentences or both. Little, if any, development of support or organizational pattern or both is apparent. Limited or inappropriate word choice frequently obscures meaning. Gross errors in sentence structure and usage may imped communication. Frequent and blatant errors may occur in the basic conventions of mechanics and usage, and commonly used words may be misspelled.

Score of 2: The writing is related to the topic but includes extraneous or loosely related material. Little evidence of an organizational pattern may be demonstrated, and the paper may lack a sense of completeness or wholeness. Development of support is inadequate of illogical. Word choice is limited, inappropriate, or vague. There is little, if any variation in sentence structure, and gross errors in sentence structure may occur. Errors in basic conventions of mechanics and usage may occur, and commonly used words may be misspelled.

Score of 3: The writing is generally focused on the topic but may include extraneous of loosely related material. An organizational pattern has been attempted, but the paper may lack a sense of completeness or wholeness. Some support is included, but development is erratic. Word choice is adequate but may be limited, predictable, or occasionally vague. There is little, if any, variation in sentence structure. Knowledge of the

conventions of mechanics and usage is usually demonstrated, and commonly used words are usually spelled correctly.

Score of 4: The writing is generally focused on the topic but may include extraneous or loosely related material. An organizational pattern is apparent, although lapses may occur. The paper exhibits some sense of completeness or wholeness. The support, including word choice, is adequate, although development may be uneven. There is little variation in sentence structure, and most sentences are complete. The paper generally follows the conventions of mechanics, usage, and spelling.

Score of 5: The writing focuses on the topic, and its organizational pattern provides for a progression of ideas, although some lapses may occur. The paper conveys a sense of completeness or wholeness. The support is ample. The writing demonstrates a mature command of the language, including precision in word choice. There is variation in sentence structure, and, with rare exceptions, sentences are complete except when fragments are used purposefully. The paper generally follows the conventions of mechanics, usage, and spelling.

Score of 6: The writing is focused, purposeful, and reflects insight into the writing situation. The paper conveys a sense of completeness and wholeness with adherence to the main idea, and its organizational pattern provides for a logical progression of ideas. The support is substantial, specific, relevant, concrete, and/or illustrative. The paper demonstrates a commitment to and an involvement with the subject, clarity in presentation of ideas, and may use creative writing strategies appropriate to the purpose of the paper. The writing demonstrates a mature command of the language (word choice) with

freshness of expression. Sentence structure is varied, and sentences are complete except when fragments are used purposefully. Few, if any, convention errors occur in mechanics, usage, and punctuation.

Unscorable (0): The score of unscorable is given to three categories of responses. First, the student response is not related to what the prompt requested the student to do. Second, the student refuses to respond or simply copies or paraphrases the prompt as a response. Finally, the writing folder is blank, the response is illegible or incomprehensible (words are arranged in such a way that no meaning is conveyed), the response is written in a foreign language, or the response contains an insufficient amount of writing to determine if the student was attempting to address the prompt.

Source: Florida Statewide Assessment Program, 1995. Florida Writes! 1995 Technical Report, pp. 11-12.

APPENDIX F
SCORING DISTRIBUTIONS FOR FLORIDA WRITES! PROMPTS

Frequency Distribution, Percentage, and Cumulative Percentage of Holistic Score Points for the Grade 8 Expository Prompt

Holistic Score Points	Frequency	Percentage	Cumulative Percentage
0.0	711	1.1	1.1
1.0	1,430	2.3	3.4
1.5	2,090	3.3	6.7
2.0	8,428	13.4	20.1
2.5	8,303	13.2	33.3
3.0	16,581	26.3	59.6
3.5	10,533	16.7	76.4
4.0	9,860	15.7	92.0
4.5	2,851	4.5	96.6
5.0	1,486	2.4	98.9
5.5	442	0.7	99.6
6.0	231	0.4	100.0
Total	62,946		

Source: Florida Statewide Assessment Program, 1995. Florida Writes! 1995 Technical Report, p. 24.

Frequency Distribution, Percentage, and Cumulative Percentage of Holistic Score Points for the Grade 8 Persuasive Prompt

Holistic Score Points	Frequency	Percentage	Cumulative Percentage
0.0	859	1.4	1.4
1.0	1,682	2.7	4.1
1.5	1,944	3.1	7.2
2.0	8,355	13.4	20.5
2.5	8,059	12.9	33.4
3.0	16,506	26.4	59.8
3.5	10,311	16.5	76.3
4.0	9,537	15.3	91.6
4.5	3,149	5.0	96.6
5.0	1,394	2.2	98.9
5.5	504	0.8	99.7
6.0	208	0.3	100.0
Total	62,508		

Source: Florida Statewide Assessment Program, 1995. Florida Writes! 1995 Technical Report, p. 25.

APPENDIX G
PARENT CONSENT FORM

Dear Parent/Guardian:

I am a doctoral student in the Foundations of Education department at the University of Florida studying under the supervision of Dr. John Kranzler, Associate Professor. I am currently conducting research that examines whether styles of thinking can impact students' performance on various types of tests. More specifically, my study will examine whether students with certain thinking styles perform better on the *Florida Writes!* assessment than on the *Iowa Test of Basic Skills*. These results may not directly help your child today, but they may benefit future students as we learn more about the characteristics or traits that influence students' performance on various tests of academic achievement.

All participants will be administered a brief measure of thinking style and cognitive ability by either myself or another qualified graduate student. The thinking style inventory requires the students to read a series of statements and decide how well each statement fits the way they typically do things at school and home. The measure of cognitive ability requires the students to select pictures that best complete a figural pattern. Both of these measures will be group-administered, and total assessment time will be approximately one and a half hours. The measures will be administered during school hours during a period that is agreed upon by your child's teacher and the school principal. With your permission, your child's most recent scores on the *Iowa Test of Basic Skills* and the *Florida Writes!* tests will be accessed and matched to their scores on the measures of thinking style and cognitive ability. After this initial matching, <u>all</u> identifying information (e.g. names) will be replaced with code numbers. Your child's identity will be kept confidential to the extent provided by law, and results will only be reported in the form of group data. Furthermore, participation in this study will not affect your child's grade in any class or placement in any programs.

You and your child have the right to withdraw consent for your child's participation at any time without consequence. Your child does not have to answer any question or item that s/he does not wish to answer. There are no known risks to the participants, and most students find the measure of thinking styles interesting and informative. No compensation is offered for participation, but group results of this study will be available in 1999 by request. If you have any questions about this research project, please contact me at (352) 392-0723 ext. 295 or my faculty supervisor, Dr. Kranzler at 392-0723 ext. 236. Questions or concerns about research participants' rights may be directed to the UFIRB office, University of Florida, Box 112250, Gainesville, FL 32611, (352) 392-0433.

Sincerely,

Briley Proctor

I have read the procedure described above. I voluntarily give my consent for my child, _____, to participate in Briley Proctor's study of cognitive styles. I also give permission for Briley Proctor to access my child's *Iowa Test of Basic Skills* and *Florida Writes!* scores from the Alachua County School Board. I have received a copy of this description.

_____ _____

Parent/Guardian Date

_____ _____

2nd Parent/Witness Date

Child's Homeroom Teacher: _____

APPENDIX H
STUDENT ASSENT SCRIPT

Hi, my name is _____ and I am a graduate student at the University. Today I will be asking all of you to complete two paper-and-pencil tasks. The first task asks you to read statements about the strategies that you use to solve problems, to carry out projects, and to make decisions. You will then be asked to decide how well each statement describes you. The second task requests that you solve a series of puzzles that are similar to analogies. More specific directions for completing each task will be given to you in just a few minutes. You will have approximately one and a half hours to complete both tasks. You may stop participating at any time, and you do not have to answer any question that you do not wish to answer. You will not in any way be penalized if you choose not to participate. Furthermore, your scores on these tasks will not affect your grades.

If you participate in this study, I will need to match your scores on the two tasks you complete today with your scores on the Iowa Test of Basic Skills and the Florida Writes! assessment. Your parent or guardian has already given me their permission to access these scores from the school board. Once I match the scores, your names will be replaced by a subject identification number in order to ensure confidentiality. No one but myself will ever see your scores on these tasks, and once I've replaced your name with a subject number, I will no longer be able to link any scores to your name.

Are there any questions? If you still agree to participate, you may begin as soon as I have distributed the tasks and given some brief directions. If you do not wish to participate you may leave now, or you may turn the papers in blank with the others.

APPENDIX I
FULL CORRELATION MATRIX

	FW	ITBS	SPM	Exec	Legis	Jud	Global	Local	Int	Ext	Lib	Conserv	Hier	Mon	Olig	Anarch
FW																
ITBS	.50**															
SPM	.24**	.51**														
Exec	.13	-.01	.05													
Legis	.08	.16*	.14	.13												
Jud	.12	.09	.05	.47**	.36**											
Global	.03	.06	.01	.20*	.38**	.32**										
Local	.16	.06	.06	.46**	.36**	.52**	.32**									
Internal	.12	.20*	.11	.00	.37**	.17*	.24**	.26**								
External	.16	.02	.12	.34**	.24**	.47**	.24**	.34**	-.19*							
Liberal	.08	.02	.19*	.22**	.51**	.56**	.42**	.51**	.35**		.36**					
Conserv	-.02	-.09	-.11	.69**	.07	.38**	.27**	.36**	-.08*		.25**	.08				
Hierarch	.18*	.09	.11	.47**	.31**	.50**	.32**	.53**	.26**		.30**	.45**	.31**			
Monarch	.08	-.09	-.06	.37**	.23**	.21**	.30**	.42**	.27**		.22**	.21**	.37**	.43**		
Oligarchic	.13	-.06	-.02	.52**	.04	.44**	.17*	.45**	-.14**		.45**	.23**	.48**	.34**	.35**	
Anarchic	.20	.03	.16*	.34**	.36**	.51**	.37**	.44**	.20*		.48**	.55**	.25**	.49**	.33**	.31**

REFERENCES

Anastasi, A. (1988). Psychological testing (6th ed.). New York: Macmillan.

Bandalos, D. (1996). Confirmatory factor analysis. In J. Stevens (Ed.), Applied multivariate statistics for the social sciences (3rd ed., pp. 389-418). Mahwah, NJ: Erlbaum.

Berry, J.W. (1976). Human ecology and cognitive style: Comparative studies in cultural and psychological adaptation. Beverly Hills, CA: Sage.

Bryant, F.B., & Yarnold, P.R. (1997). Principal-components analysis and exploratory and confirmatory factor analysis. In L.G. Grimm & P.R. Yarnold (Eds.), Reading and understanding multivariate statistics (pp. 99-136). Washington, DC: American Psychological Association.

Burke, H.R. (1958). Raven's Progressive Matrices: A review and critical evaluation. Journal of Genetic Psychology, 93, 199-228.

Cooper, S.E., & Miller, J.A. (1991). MBTI learning style-teaching style discongruencies. Educational and Psychological Measurement, 51, 699-706.

Corman, L., & Budoff, M. (1974). Factor structures of Spanish-speaking and non-Spanish-speaking children on Raven's Progressive Matrices. Educational and Psychological Measurement, 34, 977-981.

Council of Chief State School Officers (1998). Key state education policies on K-12 education: Standards, graduation, assessment, teacher licensure, time and attendance. A 50-state report August 1998. Washington, DC: Author.

CTB Macmillan/McGraw-Hill (1985). California Achievement Test. Monterey, CA: Author.

DeBello, T.C. (1989). Comparison of eleven major learning styles models. Paper presented at the National Conference of the Association for Supervision and Curriculum Development, Orlando, FL.

DuBois, P.H. (1970). A history of psychological testing. Boston: Allyn and Bacon.

Dunbar, S.B. (1987). Comparability of indirect assessments of writing skills as predictors of writing performance across demographic groups. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, DC.

Dunn, R. (1984). Learning style: State of the science. Theory Into Practice, 23, 10-19.

Dunn, R., Dunn, K., & Price, G.E. (1979). Learning Styles Inventory. Lawrence, KA: Price Systems.

Dunn, R., & Griggs, S.A. (1988). Learning styles: Quiet revolution in American secondary schools. Reston, VA: NASSP.

Dunn, R., Griggs, S.A., Olson, J., Beasley, M., & Gorman, B.S. (1995). A meta-analytic validation of the Dunn and Dunn model of learning-style preferences. The Journal of Educational Research, 88, 353-362.

Dyer, C. (1985). Otis-Lennon school ability test. In J.V. Mitchell, Jr. (Ed.), The ninth mental measurements yearbook, 2, [Test # 913]. Lincoln, NE: Buros Institute of Mental Measurements.

Engelhard, G., Walker, E.S., Gordon, B., & Gabrielson, S. (1994). Writing tasks and gender: Influences on writing quality of Black and White students. Journal of Educational Research, 87, 197-209.

Finch, F.L. (Ed.). (1991). Educational performance assessment. Chicago: Riverside.

Flanagan, D.P., Andrews, T.J., & Genshaft, J.L. (1997). The functional utility of intelligence tests with special education populations. In D.P. Flanagan, J.L. Genshaft, & P.L. Harrison (Eds.), Contemporary intellectual assessment: Theories, Tests, and Issues (pp. 457-483). New York: Guilford.

Florida Statewide Assessment Program (1995). Florida Writes! 1995 technical report. Tallahassee, FL: Department of Education.

Frederiksen, J.R., & Collins, A. (1989). A systems approach to educational testing. Educational Researcher, 18, 27-32.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. American Psychologist, 39, 193-202.

Gardner, R.W. (1953). Cognitive styles in categorizing behavior. Journal of Personality, 22, 214-233.

Gardner, R.W., Jackson, D.N., & Messick, S.J. (1960). Personality organization in cognitive controls and intellectual abilities. Psychological Issues, 2, 4.

Gardner, R.W., & Moriarty, A. (1968). Dimensions of cognitive control at preadolescence. In R. Gardner (Ed.), Personality development at preadolescence. Seattle: University of Washington Press.

Goldstein, K.M., & Blackman, S. (1978). Cognitive style: Five approaches and relevant research. New York: Wiley.

Goodenough, D.R. (1976). The role of individual differences in field dependence as a factor in learning and memory. Psychological Bulletin, 83, 675-694.

Goodenough, D.R., & Karp, S. A. (1961). Field dependence and intellectual functioning. Journal of Abnormal and Social Psychology, 63, 241-246.

Gough, H.G., & Sampson, H. (1954). The College Vocabulary Test. Berkeley, CA: Institute of Personality Assessment and Research, University of California, Berkeley.

Gregorc, A.F. (1982). Gregorc Style Delineator. Maynard, MA: Gabriel Systems.

Gregorc, A.F. (1984). Gregorc Style Delineator: Development, technical, and administrative manual. Maynard, MA: Gabriel Systems.

Gregorc, A.F., & Ward, H.B. (1977). A new definition for individual. NASSP Bulletin, February, 20-26.

Grigorenko, E.L., & Sternberg, R.J. (1993a). Set of Thinking Styles Tasks for Students. Unpublished test.

Grigorenko, E.L., & Sternberg, R.J. (1993b). Students' Thinking Styles Evaluated by Teachers. Unpublished test.

Grigorenko, E.L., & Sternberg, R.J. (1993c). Thinking Styles Questionnaire for Teachers. Unpublished test.

Grigorenko, E.L., & Sternberg, R.J. (1997). Styles of thinking, abilities, and academic performance. Exceptional Children, 63, 295-312.

Henson, K.T., & Borthwick, P. (1984). Matching styles: A historical look. Theory Into Practice, 23, 3-9.

Herrnstein, R.J., & Murray, C. (1994). The bell curve: Intelligence and class structure in American life. New York: The Free Press.

Hoover, H.D., Hieronymus, A.N., Frisbie, D.A., Dunbar, S.B., Oberley, K.R., Cantor, N.K., Bray, G.B., Lewis, J.C., & Qualls-Payne, A.L. (1993). Iowa Test of Basic Skills: Interpretive guide for teachers and counselors. Chicago: Riverside.

Hyde, J.S., & Linn, M.C. (1988). Gender differences in verbal ability: A meta-analysis. Psychological Bulletin, 104, 53-69.

Irvine, J.J. (1990). Black students and school failure: Policies, practices, and prescriptions. New York: Guilford.

Jensen, A. (1980). Bias in mental testing. New York: Free Press.

Jung, C.G. (1971). Psychological types (H.G. Baynes, Trans. revised by R.F.C. Hull) Princeton, NJ: Princeton University Press. (Original work published in 1921).

Kagan J., Rosman B., Day D., Albert, J., & Phillips, W. (1964). Information processing in the child: Significance of analytic and reflective attitudes. Psychological Monographs, 78, 1.

Kaiser, H.F. (1958). The varimax criterion for analysis rotation in factor analysis. Psychometrika,23, 187-200.

Kamphaus, R.W., & Frick, P.J. (1996). Clinical assessment of child and adolescent personality and behavior. Boston: Allyn & Bacon.

Kavale, K.A., & Forness, S.R. (1987). Substance over style: Assessing the efficacy of modality testing and teaching. Exceptional Children, 54, 228-239.

Keefe, J.W. (1987). Learning style: Theory & practice. Reston, VA: National Association of Secondary School Principals.

Kirby, P. (1979). Cognitive styles, learning style and transfer skill acquisition. Information Series No. 198, National Center for Research in Vocational Education. Columbus, OH: Ohio State University.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. Educational Measurement: Issues and Practice, 5-15.

LeMahieu, P.G., Gitomer, D.H, & Eresh, J.T. (1995). Portfolios in large-scale assessment: Difficult but not impossible. Educational Measurement: Issues and Practice, 11-28.

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1981). Complex, performance-based assessment: Expectations and validation criteria. Educational Researcher, 20, 15-21.

Mamchur, C. (1996). A teacher's guide to cognitive type theory & learning style. Alexandria, VA: Association for Supervision and Curriculum Development.

McCaulley, M.H. (1990). The Myers-Briggs Type Indicator: A measure for individuals and groups. Measurement and Evaluation in Counseling and Development, 22, 181-195.

Messer, S.B. (1976). Reflection-impulsivity: A review. Psychological Bulletin, 83, 1026-1052.

Mueller, M. (1991). Introduction. In F.L. Finch (Ed.), Educational performance assessment (pp. vii-x). Chicago: Riverside.

Myers, I.B., & McCaulley, M.H. (1985). Manual: A guide to the development and use of the Myers-Briggs Type Indicator. Palo Alto, CA: Consulting Psychological Press.

National Association of Secondary School Principals (1979). Student learning styles-Diagnosing and prescribing programs. Reston, VA: NASSP.

Neisser, U., Boodoo, G., Bouchard, T.J., Boykin, A.W., Brody, N., Ceci, S.J., Halpern, D.F., Loehlin, J.C., Perlin, R., Sternberg, R.J., & Urbina, S. (1996). Intelligence: Knowns and Unknowns. American Psychologist, 51, 77-100.

Oakes, J. (1985). Keeping track: How schools structure inequality. New Haven, CT: Yale University Press.

Oakland, T. (1985). Otis-Lennon school ability test. In J.V. Mitchell, Jr. (Ed.), The ninth mental measurements yearbook, 2, [Test # 913]. Lincoln, NE: Buros Institute of Mental Measurements.

Pettigrew, T.F. (1958). The measurement and correlates of category width as a cognitive variable. Journal of Personality, 26, 532-544.

Pittenger, D.J. (1993). The utility of the Myers Briggs Type Indicator. Review of Educational Research, 63, 467-488.

Raskin, E. (1985). Counseling implications of field dependence-independence in an educational setting. In M. Bertini, L. Pizzamiglio, & S. Wapner (Eds.), Field dependence in psychological theory, research, and application: Two symposia in memory of Herman A. Witkin (pp. 107-113). Hillsdale, NJ: Erlbaum.

Raven, J.C. (1938). Guide to using Progressive Matrices. London: Lewis.

Raven, J.C., Court, J.H., & Raven, J. (1983). Manual for Raven's Progressive Matrices and Vocabulary Scales (Section 3)-Standard Progressive Matrices (1983 edition). London: Lewis.

Resnick, L.B., & Resnick, D.P. (1991). Assessing the thinking curriculum: New tools for educational reform. In B.R. Gifford & M.C. O'Connor (Eds.), Changing assessments: Alternative views of aptitude, achievement, and instruction (pp. 37-75). Boston: Kluwer.

Riverside Publishing Company (1994). Riverside 2000 integrated assessment program technical summary I. Chicago: Author.

Rollins, T.J. (1990). Analysis of theoretical relationship between learning styles of students and their preferences for learning activities. Journal of Agricultural Education, 31, 64-70.

Rorschach, H. (1921). Psychodiagnostik. Bern: Bircher.

Sattler, J.M. (1988). Assessment of children (Rev. ed.). San Diego: Author.

Sewall, T.J. (1986). The measurement of learning style: A critique of four assessment tools. Green Bay, WI: Wisconsin University. (ERIC Document Reproduction Service No. ED 267 247).

Sternberg, R.J. (1988). The triarchic mind: A new theory of human intelligence. New York: Viking.

Sternberg, R.J. (1990). Thinking styles: Keys to understanding student performance. Phi Delta Kappan, 366-370.

Sternberg, R.J. (1993). Sternberg Triarchic Abilities Test. Unpublished test.

Sternberg, R.J. (1994a). Allowing for thinking styles. Educational Leadership, 36-40.

Sternberg, R.J. (1994b). Thinking styles: Theory and assessment at the interface between intelligence and personality. In R.J. Sternberg and P. Ruzgis (Eds.), Intelligence and personality. New York: Cambridge University Press.

Sternberg, R.J. (1997a). Are cognitive styles still in style? American Psychologist, 52, 700-712.

Sternberg, R.J. (1997b). Thinking styles. New York: Cambridge University Press.

Sternberg, R.J., & Grigorenko, E.L. (1993). Thinking styles and the gifted. Roeper Review, 16, 122-130.

Sternberg, R.J., & Grigorenko, E.L. (1995). Styles of thinking in the school. European Journal for High Ability, 6, 201-219.

Sternberg, R.J., & Wagner, R.K. (1991). MSG Thinking Styles Inventory manual. Unpublished test manual.

Thorne, A., & Gough, H. (1991). Portraits of type: An MBTI research compendium. Palo Alto, CA: CPP Books.

U.S. Department of Education. National Center for Education Statistics. Digest of Education Statistics, 1996, NCES96-133, by Thomas D. Snyder. Production Manager, Charlene M. Hoffman. Program Analyst, Claire M. Geddes. Washington, DC: Author.

Vaidya, S., & Chansky, N. (1980). Cognitive development and cognitive style in mathematics achievement. Journal of Educational Psychology, 72, 326-330.

Weissenberg, P., & Gruenfeld, L.W. (1966). Relationships among leadership dimensions and cognitive style. Journal of Applied Psychology, 50, 392-395.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. Phi Delta Kappan, 705-713.

Wisconsin Education Association Council (1996). Educational issues series: Performance assessment. Retrieved November 18, 1997 from the World Wide Web: http://weac.org/resource/may96/perform.htm

Witkin, H.A., Dyk, R.B., Faterson, H.F., Goodenough, D.R., & Karp, S.A. (1974). Psychological differentiation: Studies in development. New York: Wiley.

Witkin, H.A., & Goodenough, D.R. (1977). Field dependence and interpersonal behavior. Psychological Bulletin, 84, 661-689.

Witkin, H.A., & Goodenough, D.R. (1981). Cognitive styles: Essence and origins. Psychological Issues Monograph 51.

Wolf, D., Bixby, J., Glenn, J., III, and Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. Review of Research in Education, 17, 31-74.

Zhang, L.F., & Sachs, J. (1997). Assessing thinking styles in the theory of mental self-government: A Hong Kong validity study. Psychological Reports, 81, 914-928.

BIOGRAPHICAL SKETCH

Briley Elizabeth Proctor was born and raised in Tallahassee, Florida, where she graduated from Leon High School in 1987. In 1992, Briley received her Bachelor of Arts degree from the University of Colorado, Boulder, with a major in psychology.

In 1995, Briley entered the APA-accredited school psychology doctoral program at the University of Florida in Gainesville, Florida. She worked for three years as a graduate assistant, primarily through a grant funded by the Genentech Foundation for Growth and Development. This work experience provided training in psychoeducational assessment, use of reaction and inspection time instrumentation, data analysis, and data synthesis. In addition, she taught undergraduate courses in the College of Education and served on numerous departmental committees.

Briley completed her Master of Arts degree in August 1997. During the 1998-1999 academic year, she completed her pre-doctoral internship at the Munroe-Meyer Institute in Omaha, Nebraska, which is part of the APA-approved Nebraska Consortium in Professional Psychology. As an intern, Briley provided psychological services to a wide variety of children and their families, many with developmental disabilities. In addition, Briley was a consultant to the BoysTown multidisciplinary team serving deaf and hard-of-hearing children, and to several local Head Start centers. During the internship year, Briley

continued to publish articles in peer-reviewed professional journals and to present at several state and national conferences.

Briley will receive her Specialist in Education and Doctor of Philosophy degrees in August 1999. Her specialization is educational measurement. Briley has accepted a position as assistant professor in the school psychology program at Florida State University, Tallahassee, Florida, where she will focus on graduate teaching and training responsibilities and will continue her research in the areas of academic and behavioral interventions.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John H. Kranzler, Chair
Associate Professor of
Foundations of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Linda M. Crocker
Professor of Foundations of
Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Edward W. Wolfe
Assistant Professor of
Foundations of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Jennifer M. Asmus
Assistant Professor of
Foundations of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Mary K. Dykes
Professor of Special
Education

This dissertation was submitted to the Graduate Faculty of the College of Education and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August 1999

Chairman, Foundations of
Education

Dean, College of Education

Dean, Graduate School